

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:55:45 ; Search time 29 Seconds
(without alignments)
113.681 Million cell updates/sec

Title: US-10-006-223-1

Perfect score: 82
Sequence: 1 VVGGRRAAGGEPFMV 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

1 number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	93.9	259	Q54168	streptomyce
2	62	75.6	255	Q9Y7A9	metarhizium
3	62	75.6	261	Q00344	cochliobolu
4	61	74.4	311	Q8WZM5	trichoderma
5	60	73.2	254	Q01136	metarhizium
6	60	73.2	256	Q9Y842	metarhizium
7	59	72.0	265	Q74596	phaeosphaer
8	56	68.3	360	Q9W1X6	drosofila
9	55	67.1	281	Q76898	drosofila
10	54	65.9	242	Q96687	lumbricus b
11	54	65.9	251	Q9VXC9	drosofila
12	54	65.9	294	Q9V7R3	drosofila
13	54	65.9	974	Q90WD8	bufo japoni
14	53	64.6	268	Q46151	pacifastacu
15	53	64.6	855	Q9JJI7	rattus norv
16	52	63.4	155	Q95VT4	homarus ame

17	52	63.4	548	16	Q9KRJ1	Q9krj1 vibrio chol
18	51	62.2	262	5	Q9VRF9	Q9vrf9 drosophila
19	51	62.2	405	5	Q44331	Q44331 manduca sex
20	51	62.2	422	4	Q8WVC1	Q8wvc1 homo sapien
21	51	62.2	638	11	Q8R0P5	Q8r0p5 mus musculu
22	50	61.0	243	5	O01310	O01310 botryllus s
23	50	61.0	271	5	Q9VRS6	Q9vrs6 drosophila
24	50	61.0	281	5	Q8SYB5	Q8syb5 drosophila
25	50	61.0	1379	5	Q9V4N6	Q9v4n6 drosophila
26	49	59.8	272	5	Q9XYV6	Q9xyv6 rhizopertha
27	49	59.8	275	6	Q28609	Q28609 oryctolagus
28	49	59.8	286	5	Q96900	Q96900 scolopendra
29	49	59.8	542	5	Q917J3	Q917j3 drosophila
30	49	59.8	613	13	Q03711	Q03711 xenopus lae
31	49	59.8	845	13	Q9DGR1	Q9dgr1 xenopus lae
32	48	58.5	25	5	P81802	P81802 eisenia foe
33	48	58.5	213	5	Q9VEN0	Q9ven0 drosophila
34	48	58.5	265	5	Q9VVT3	Q9vvt3 drosophila
35	48	58.5	433	5	Q8T3A2	Q8t3a2 ciona intes
36	47	57.3	223	5	Q9XY58	Q9xy58 ctenocephal
37	47	57.3	232	5	Q9TY45	Q9ty45 penaeus van
38	47	57.3	263	5	Q9TY16	Q9ty16 penaeus van
39	47	57.3	266	5	Q27761	Q27761 penaeus van
40	47	57.3	277	5	O96899	O96899 scolopendra
41	47	57.3	279	11	Q9QZ74	Q9qz74 rattus norv
42	47	57.3	377	6	P79343	P79343 bos taurus
43	47	57.3	417	11	Q8VHJ4	Q8vhj4 rattus norv
44	47	57.3	819	16	Q9A500	Q9a500 caulobacter
45	46	56.1	40	6	Q9TRU3	Q9tru3 oryctolagus

ALIGNMENTS

RESULT 1

Q54168 ID Q54168 PRELIMINARY; PRT; 259 AA.
AC Q54168;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Trypsinogen.
OS Streptomyces fradiae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1906;
RN [1]_TaxID=1906;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 14544;
RA Katoh T., Kikuchi N., Nagata K., Yoshida N.;
RT "Cloning and expression of Trypsin-Like enzyme from Streptomyces
fradiae for comparative analysis of functional regions of Streptomyces
and Mammalian Trypsins.";
RL J. Ferment. Bioeng. 80:440-445(1995).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; D16687; BAA04089.1; --
DR HSSP; P00775; ISGT.
DR MEROPS; S01.101; --
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT CHAIN 39 259 TRYPSIN.
SQ SEQUENCE 259 AA; 26702 MW; 8B6C8496AB840FCF CRC64;

Query Match 93.9%; Score 77; DB 2; Length 259;
Best Local Similarity 93.8%; Pred. No. 1.8e-05;

Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16
 |||||:|||||:
 Db 39 VVGCTRAAQGEFFPMV 54

RESULT 2

Q9Y7A9 PRELIMINARY; PRT; 255 AA.
 AC Q9Y7A9;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 Trypsin-related protease.
 TRY2.
 OS Metarhizium anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
 QY NCBI_TaxID=5530;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=ME1;
 RA Screen S.E., St Leger R.J.;
 RT "Isolation of multiple protease genes from the entomopathogenic fungus
 Metarhizium anisopliae.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.

DR EMBL; AF130865; AAD29675.1; -.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.103; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease.
 KW Hydrolase; Protease; Serine protease.
 SQ SEQUENCE 255 AA; 26289 MW; 88DD979ED300E4B7 CRC64;

Query Match 75.6%; Score 62; DB 3; Length 255;
 Best Local Similarity 68.8%; Pred. No. 0.0078;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16
 :|||:|||||:
 30 IVGGEAAQGEFFYIV 45

RESULT 3

ID Q00344 PRELIMINARY; PRT; 261 AA.
 AC Q00344;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE ALP1.
 GN ALP1.
 OS Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 QY NCBI_TaxID=5017;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=SB11;
 RX MEDLINE=96212997; PubMed=8634479;
 RA Murphy J.M., Walton J.D.;
 RT "Three extracellular proteases from Cochliobolus carbonum: cloning and
 targeted disruption of ALP1.";
 RL Mol. Plant Microbe Interact. 9:290-297(1996).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; U39500; AAB03851.1; -.
 DR HSP; P00763; IDPO.
 DR MEROPS; S01.103; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 261 AA; 26052 MW; FD4B4A02BF44BDE1 CRC64;

Query Match 75.6%; Score 62; DB 3; Length 261;
 Best Local Similarity 68.8%; Pred. No. 0.008;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16
 :|||:|||||:
 Db 31 IVGTTAAAGEFFIV 46

RESULT 4

Q8WZM5 PRELIMINARY; PRT; 311 AA.
 ID Q8WZM5
 AC Q8WZM5;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Protease P27 (Fragment).
 GN P27.
 OS Trichoderma harzianum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; mitosporic Hypocreales; Trichoderma.
 QY NCBI_TaxID=5544;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN=CECT 2413;
 RA Suarez M.B.;
 RL Thesis (1999), University of Salamanca, Salamanca, SPAIN.
 DR EMBL; AJ249721; CAC80694.1; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser.protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 311 AA; 31200 MW; B31CC1493DB2263F CRC64;

Query Match 74.4%; Score 61; DB 3; Length 311;
 Best Local Similarity 68.8%; Pred. No. 0.015;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16
 :|||:|||||:
 Db 83 IVGTTAAAGEFFIV 98

RESULT 5

Q01136 PRELIMINARY; PRT; 254 AA.
 ID Q01136
 AC Q01136;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Trypsin-like protease 1 precursor.
 GN TYR1.

OS Metarhizium anisopliae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
 OX NCBI_TaxID=5530;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MEL;
 RX MEDLINE=96105219; PubMed=8529882;
 RA Smithson S.L., Paterson I.C., Bailey A.M., Screen S.E., Hunt B.A.,
 RA Cobb B.D., Cooper R.M., Charnley A.K., Clarkson J.M.;
 RT "Cloning and characterisation of a gene encoding a cuticle-degrading
 RT protease from the insect pathogenic fungus Metarhizium anisopliae";
 RL Gene 166:161-165(1995).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; X78875; CAA55477.1; -.
 DR HSSP; P35049; ITRY.
 DR MEROPS; S01.103; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease; Signal.
 F1 SIGNAL 1 21 POTENTIAL.
 F1 SEQUENCE 254 AA; 26117 MW; 0ABD96A5C52DBA54 CRC64;
 SQ
 Query Match 73.2%; Score 60; DB 3; Length 254;
 Best Local Similarity 68.8%; Pred. No. 0.018;
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGGTAAAGCEFFPMV 16
 :|||:|||||:
 Db 30 IVGGSAAAGCEFFPFI 45
 RESULT 6
 QY842
 ID QY842 PRELIMINARY; PRT; 256 AA.
 AC QY842;
 DT 01-NOV-1999 (TREMELrel. 12, Created)
 DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
 DE Trypsin-related protease precursor.
 TR11
 Metarhizium anisopliae.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreales; Clavicipitaceae; mitosporic Clavicipitaceae; Metarhizium.
 OX NCBI_TaxID=5530;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=MEL;
 RX Screen S.E., StLeger R.J.;
 RT "Isolation of multiple protease genes from the entomopathogenic fungus
 RT Metarhizium anisopliae";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AJ242736; CAB44652.1; -.
 DR HSSP; P35049; ITRY.
 DR MEROPS; S01.103; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease; Signal.
 KW

FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 30 256 TRYPSIN-RELATED PROTEASE.
 SQ SEQUENCE 256 AA; 26201 MW; 34696608745CB982 CRC64;
 Query Match 73.2%; Score 60; DB 3; Length 256;
 Best Local Similarity 68.8%; Pred. No. 0.018;
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGGTAAAGCEFFPMV 16
 :|||:|||||:
 Db 30 IVGGSAAAGCEFFPFI 45
 RESULT 7
 QY842
 ID QY842 PRELIMINARY; PRT; 265 AA.
 AC QY842;
 DT 01-NOV-1998 (TREMELrel. 08, Created)
 DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
 DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
 DE Trypsin-like protease.
 GN SNPI.
 OS Phaeosphaeria nodorum (Septoria nodorum).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
 OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
 OC Phaeosphaeriaceae; Phaeosphaeria.
 OX NCBI_TaxID=13684;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=LAW95;
 RX MEDLINE=20255037; PubMed=10796020;
 RA Carlile A.J., Bindschedler L.V., Bailey A.M., Bowyer P.,
 RA Clarkson J.M., Cooper R.M.;
 RT "Characterization of SNPI, a cell wall-degrading trypsin, produced
 RT during infection by Stagonospora nodorum";
 RL Mol. Plant Microbe Interact. 13:538-550(2000).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AF092435; AAC61777.1; -.
 DR HSSP; P35049; ITRY.
 DR MEROPS; S01.103; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Protease; Serine protease.
 KW SEQUENCE 265 AA; 26296 MW; 3B9DB00640DA4FEA CRC64;
 SQ
 Query Match 72.0%; Score 59; DB 3; Length 265;
 Best Local Similarity 62.5%; Pred. No. 0.028;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGGTAAAGCEFFPMV 16
 :|||:|||||:
 Db 37 IVGGTTASAGDFPFI 52
 RESULT 8
 QY842
 ID QY842 PRELIMINARY; PRT; 360 AA.
 AC QY842;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE CG3700 protein (GH06673p).
 GN CG3700.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Query Match	69.3%	Score 56;	DB 5;	Length 360;
Best Local Similarity	66.7%	Pred. NO.	0.13;	
Matches 10;	Conservative	3;	Mismatches	2;
			Indels	0;
			Gaps	0;


```
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AE003806; AAF57983.1; -.
DR HSP; P20160; IA7S.
DR FlyBase; FBgn0034139; CG4927.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW SEQUENCE 294 AA; 32482 MW; 4867A6C355BB3C51 CRC64;
SQ SEQUENCE 294 AA; 32482 MW; 4867A6C355BB3C51 CRC64;
Query Match 65.9%; Score 54; DB 5; Length 294;
Best Local Similarity 66.7%; Pred. No. 0.23;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGGTAAAGGEFFPM 15
DB 37 IVGGAKAAGREFFPM 51
RESULT 13
Q90WD8 Q90WD8 PRELIMINARY; PRT; 974 AA.
01-DEC-2001 (TrEMBLrel. 19, Created)
01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Oviductin.
OS Bufo japonicus (Japanese toad).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae; Bufo.
NCBI_TaxID=8387;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVIDUCTAL PARS RECTA;
RA Hiyoishi M., Takamune K., Mita K., Kubo H., Sugimoto Y., Katagiri C.;
RT "Oviductin, the oviductal protease that mediates gamete interaction by
RT affecting the vitelline envelope in Bufo japonicus: Its molecular
RT cloning and analyses of expression and post-translational
RT activation."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB070367; BAB63372.1; -.
DR MEROPS; S01.240; -.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00089; trypsin; 2.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 2.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 974 AA; 107647 MW; F197054A70465553 CRC64;
```

```
Query Match 65.9%; Score 54; DB 13; Length 974;
Best Local Similarity 62.5%; Pred. No. 0.85;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGGTAAAGGEFFPMV 16
DB 50 IVGTSVAVGESPMV 65
RESULT 14
O46151 O46151 PRELIMINARY; PRT; 268 AA.
AC O46151;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4) (Fragment).
OS Pacifastacus leniusculus (Signal crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Astacidae; Pacifastacus.
NCBI_TaxID=6720;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEPATOPANCREAS;
RA Hernandez-Cortes M.P., Cerenius L., Garcia-Carrero F.L., Soderhall K.;
RT "Trypsin from Pacifastacus leniusculus Hepatopancreas: Purification
RT and cDNA Cloning of the Synthesized Zymogen."
RL Biol. Chem. 380:499-501(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ222658; CAA10915.1; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.122; -.
DR InterPro; IPR001969; Asprotease_site.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1 17 POTENTIAL.
FT SIGNAL <1 17 TRYPSIN.
FT CHAIN 31 266 TRYPSIN.
SQ SEQUENCE 268 AA; 28482 MW; 9080F9736412FEF5 CRC64;
Query Match 64.6%; Score 53; DB 5; Length 268;
Best Local Similarity 56.2%; Pred. No. 0.32;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGGTAAAGGEFFPMV 16
DB 32 IVGTDASLGEFFYQL 47
RESULT 15
Q9JJI7 Q9JJI7 PRELIMINARY; PRT; 855 AA.
AC Q9JJI7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Membrane bound serine protease (Membrane bound arginine specific
DE serine protease).
GN MBSP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=JEJUNUM;
RA Tezuka S.;
RT "A membrane bound serine protease expressed in rat small intestine.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=DUODENUM;
RA Inoue H., Takahashi K., Kishi K.;
RT "membrane-bound arginine specific serine protease.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB037898; BAB03502.1; -.
DR EMBL; AB049189; BAB13765.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.302; -.
DR InterPro; IPR002106; AACRNA_ligaseI.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000859; CUB domain.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00431; CUB; 2.
DR Pfam; PF00057; ldl_recept_a; 4.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00042; CUB; 2.
DR SMART; SM00192; LDLa; 3.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR PROSITE; PS01180; CUB; 2.
DR PROSITE; PS01209; LDLRA_1; 2.
DR PROSITE; PS00068; LDLRA_2; 4.
DR PROSITE; PS00240; TRYPsin DOM; 1.
DR PROSITE; PS00134; TRYPsin HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin SER; 1.
KW Glycoprotein; Hydrolase; Protease; Serine protease.
FT VARIANT 665 665
SQ SEQUENCE 855 AA; 94955 MW; 35806B7ECF6CF03D CRC64;

Query Match 64.6%; Score 53; DB 11; Length 855;
Best Local Similarity 62.5%; Pred. No. 1.1;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 VVGGTAAQGEFFPMV 16
|||||:|:|:|
Db 615 VVGGTNADEGEWPMQV 630
```

rch completed: April 11, 2003, 17:58:07
time : 31 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:55:20 ; Search time 11 Seconds
(without alignments)
60.329 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVGGRRAAGBFFFW 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

1 number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	259	1 TRYP_STRGR	P00775 streptomyce
2	62	75.6	268	1 TRYP_STRGA	Q54179 streptomyce
3	59	72.0	248	1 TRYP_FUSOX	P35049 fusarium ox
4	58	70.7	20	1 TRYL_STRX	P80420 streptomyce
5	53	64.6	855	1 ST14_MOUSE	P56677 mus musculu
6	52	63.4	216	1 CTR2_VESOR	P00768 vespa orien
7	52	63.4	237	1 TRYP_ASTFL	P00765 astacus flu
8	52	63.4	254	1 CTRL_HALRU	P35003 haliois ru
9	52	63.4	274	1 FA9_SHEEP	P16291 ovis aries
10	51	62.2	638	1 KAL_MOUSE	P26262 mus musculu
11	51	62.2	855	1 ST14_HUMAN	Q9Y5Y6 homo sapien
12	50	61.0	32	1 TRYP_PENMO	P35050 penaeus mon
13	50	61.0	218	1 CTR2_VESCR	P00769 vespa crabr
14	49	59.8	219	1 CAP7_PIG	P80015 sus scrofa
15	49	59.8	431	1 ACRO_RABIT	P48038 oryctolagus
16	49	59.8	625	1 FAL1_HUMAN	P03951 homo sapien
17	48	58.5	20	1 COG3_PARCM	P20734 paralithode
18	48	58.5	271	1 S24D_ANOGA	Q17004 anopheles g
19	48	58.5	638	1 KAL_HUMAN	P03952 homo sapien
20	48	58.5	638	1 KAL_RAT	P14272 rattus norv
21	47	57.3	60	1 ACRO_CAPIH	P10626 capra hircu
22	47	57.3	416	1 FA9_BOVIN	P00741 bos taurus
23	46	56.1	271	1 FA9_PIG	P16293 sus scrofa
24	46	56.1	275	1 FA9_RABIT	P16292 oryctolagus
25	46	56.1	705	1 CIR_HUMAN	P00736 homo sapien
26	45	54.9	20	1 COG2_CHIOP	P34154 chionocete
27	45	54.9	20	1 COG3_CHIOP	P34155 chionocete
28	45	54.9	20	1 COG4_PARCM	P20732 paralithode
29	45	54.9	20	1 COG5_PARCM	P20733 paralithode
30	45	54.9	267	1 ELNE_HUMAN	P08246 homo sapien
31	45	54.9	285	1 FA9_CAVPO	P16295 cavia porce
32	45	54.9	418	1 HAT1_HUMAN	O60235 homo sapien
33	45	54.9	422	1 DES1_HUMAN	Q9UL52 homo sapien

RESULT 1

ID	TRYP_STRGR	STANDARD;	PRT;	259 AA.
AC	P00775;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Trypsin precursor (EC 3.4.21.4) (SGT).			
GN	SPRT.			
OS	Streptomyces griseus.			
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;			
OC	Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1911;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 10137;			
RX	MEDLINE=92095977; PubMed=1755852;			
RA	Kim J.C., Cha S.H., Jeong S.T., Oh S.K., Byun S.M.;			
RT	"Molecular cloning and nucleotide sequence of Streptomyces griseus			
RT	trypsin gene."			
RL	Biochem. Biophys. Res. Commun. 181:707-713(1991).			
RN	[2]			
RP	SEQUENCE OF 37-259.			
RX	MEDLINE=75127940; PubMed=804314;			
RA	Olafson R.W., Jurasek L., Carpenter M.R., Smillie L.B.;			
RT	"Amino acid sequence of Streptomyces griseus trypsin. Cyanogen			
RT	bromide fragments and complete sequence."			
RL	Biochemistry 14:1168-1177(1975).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).			
RX	MEDLINE=88286735; PubMed=3135412;			
RA	Read R.J., James M.N.G.;			
RT	"Refined crystal structure of Streptomyces griseus trypsin at 1.7-A			
RT	resolution."			
RL	J. Mol. Biol. 200:523-551(1988).			
CC	-1- CATALYTIC ACTIVITY: Preferential cleavage: Arg- -Xaa, Lys- -Xaa.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sb-sib.ch).			
CC	-----			
CC	EMBL; M64471; AAA26820.1; ALT_SEQ.			
DR	PIR; A00962; TRSMG			
DR	PIR; JQ1302; JQ1302.			
DR	PDB; ISGT; 16-JUL-88.			
DR	MEROPS; S01.101; -			
DR	InterPro; IPR001254; Chymotrypsin.			
DR	InterPro; IPR001254; Ser protease_Try.			
DR	Pfam; PF00089; trypsin.1.			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; Tryp_Spc; 1.			

P19540 canis famil
P00740 homo sapien
P39675 dermatophag
P09370 euroglyphus
P05156 homo sapien
P98073 homo sapien
P20160 homo sapien
P08001 sus scrofa
P10323 homo sapien
Q9J1q8 mus musculu
Q9WUW3 rattus norv
O00187 homo sapien

```
DR PROSITE; PS50240; TRYPSIN DOM; 1.
DR PROSITE; PS00134; TRYPSIN HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal; 3D-structure.
FT SIGNAL 1 32
FT PROPEP 33 36 ACTIVATION PEPTIDE.
FT CHAIN 37 259 TRYPSIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM.
FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
FT ACT_SITE 208 208 CHARGE RELAY SYSTEM.
FT DISULFID 58 74
FT DISULFID 177 192
FT DISULFID 204 233
FT SITE 202 202 REQUIRED FOR SPECIFICITY.
FT CONFLICT 95 96 MISSING (IN REF. 2).
FT STRAND 38 38
FT TURN 39 39
FT STRAND 41 42
FT TURN 45 46
FT TURN 49 50
FT STRAND 51 54
FT TURN 55 57
FT STRAND 58 64
FT TURN 65 66
FT STRAND 67 70
FT HELIX 72 74
FT STRAND 79 80
FT STRAND 85 88
FT STRAND 92 92
FT TURN 93 94
FT TURN 96 97
FT STRAND 99 108
FT TURN 110 111
FT STRAND 120 124
FT STRAND 134 135
FT TURN 140 141
FT STRAND 145 150
FT TURN 156 157
FT STRAND 163 163
FT STRAND 165 172
FT STRAND 174 184
FT HELIX 187 189
FT TURN 190 193
FT TURN 196 198
FT STRAND 202 202
FT TURN 205 206
FT TURN 208 209
FT STRAND 211 215
FT TURN 217 218
FT STRAND 221 229
FT TURN 236 237
FT STRAND 240 244
FT HELIX 245 257
FT TURN 258 258
SQ SEQUENCE 259 AA; 26776 MW; 050233AFF1F64823 CRC64;

Query Match 100.0%; Score 82; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGGTAAAGQGFPPMV 16
DB 37 VVGGTAAAGQGFPPMV 52

RESULT 2
ID TRYP STRGA STANDARD; PRT; 268 AA.
AC Q54179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin-like protease precursor (EC 3.4.21.-).

Query Match 75.6%; Score 62; DB 1; Length 268;
Best Local Similarity 75.0%; Pred. No. 0.0011;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGQGFPPMV 16
DB 46 VVGKPAAGQGFPPMV 61

RESULT 3
ID TRYP FUSOX STANDARD; PRT; 248 AA.
AC P35049;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Fusarium oxysporum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; mitosporic Hypocreales; Fusarium.
OX NCBI_TaxID=5507;
RN [1]
RP SEQUENCE FROM N.A., AND X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=93324511; PubMed=8332590;
RA Rypniewski W.R., Hastrup S., Betzel C., Dauter M., Dauter Z.,
RA Papendorf G., Branner S., Wilson K.S.;
RT "The sequence and X-ray structure of the trypsin from Fusarium
```

RT oxyphorum.";
RL Protein Eng. 6:341-348(1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.55 ANGSTROMS).
RA Rypniewski W.R., Dammann C., von der Osten C., Dauter M.,
RA Wilson K.S.;
RT "Structure of inhibited trypsin from *Fusarium oxysporum* at 1.55 A.";
RL Acta Crystallogr. D 51:73-84(1995).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S63827; AAB27568.1; -;
CC PDB; 1TRY; 01-JAN-96.
CC MEROPS; S01.103; -;
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS0240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydroxylase; Serine protease; Zymogen; signal; 3D-structure.
FT SIGNAL 1 17
FT PROPEP 18 25
FT CHAIN 26 248
FT ACT_SITE 65 65
FT ACT_SITE 108 108
FT ACT_SITE 204 204
FT DISULFID 50 66
FT DISULFID 174 189
FT DISULFID 200 225
FT SITE 198 198
FT SITE 198 198
SQ SEQUENCE 248 AA; 24576 MW; 1A0EBA88C3E70294 CRC64;
Query Match 72.0%; Score 59; DB 1; Length 248;
Best Local Similarity 62.5%; Pred. No. 0.0034;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db 1 VVGTTAAAGQGPFFMV 16
:|||||:|:|:|:
25 IVGTSASAGDPFFIV 40

ULT 4
L_STREX STANDARD; PRT; 20 AA.
AC P80420;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin-like protease (SC 3.4.21.-) (fragment).
OS Streptomyces exfoliatus (Streptomyces hydroganans),
OC Bacteria; Actinobacteria; Actinobacteria (Class); Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetes.
CC NCBI_TaxID=1905;
RN [1]
RP SEQUENCE.
RC STRAIN=SMF13;
RC MEDLINE=95291424; PubMed=7773379;
RA Kim I.S., Lee K.J.;
RT "Physiological roles of leupeptin and extracellular proteases in
RT mycelium development of *Streptomyces exfoliatus* SMF13.";
RL Microbiology 141:1017-1025(1995).

CC -!- FUNCTION: INVOLVED IN MYCELIUM DIFFERENTIATION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2A.
CC MEROPS; S01.101; -;
CC InterPro; IPR001254; Ser_protease_Try.
CC PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
CC PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
CC Hydroxylase; Serine protease.
FT NON_TER 20
SQ SEQUENCE 20 AA; 2129 MW; 4568F56D0E7393AP CRC64;
Query Match 70.7%; Score 58; DB 1; Length 20;
Best Local Similarity 84.6%; Pred. No. 0.00037;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 VGGTAAAGQGPFF 14
|||:|:|:|:
DB 2 VGGTAAAGQGNFF 14
|||:|:|:|:
RESULT 5
ST14_MOUSE
ID ST14_MOUSE STANDARD; PRT; 855 AA.
AC P56677;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suppressor of tumorigenicity 14 (BC 3.4.21.-) (Epithin).
GN ST14 OR PRSS14.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RX MEDLINE=99216440; PubMed=10199918;
RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
RA Schwartz R.H.;
RT "Cloning and chromosomal mapping of a gene isolated from thymic
RT stromal cells encoding a new mouse type II membrane serine protease,
RT epithin, containing four LDL receptor modules and two CUB domains.";
RL Immunogenetics 49:420-428(1999).
RN [2]
RP REVISIONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
RC STRAIN=C.B.17SCID; TISSUE=Thymus;
RA Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
CC TESTIS, AND BRAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF042822; AAD02230.3; -;
CC EMBL; BC005496; AAH05496.1; -;
CC HSP; P20231; IAAO.
CC MEROPS; S01.302; -;
CC MGD; MGI:1338881; St14.

```
DR InterPro: IPR000859; CUB domain.
DR InterPro: IPR002172; LDL receptor A.
DR InterPro: IPR001254; Ser protease_Try.
DR Pfam: PF00057; ldl_recept_a; 4.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00431; CUB; 2.
DR SMART: SM00042; CUB; 2.
DR SMART: SM00192; LDLA; 3.
DR PROSITE: PS01180; CUB; 2.
DR PROSITE: PS01209; LDLRA_1; 2.
DR PROSITE: PS00688; LDLRA_2; 4.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
DR Transmembrane; Repeat.
FT DOMAIN 1 55
FT TRANSMEM 56 77
DR DOMAIN 77 855
DR DOMAIN 214 331
DR DOMAIN 340 444
DR DOMAIN 451 488
DR DOMAIN 489 522
DR DOMAIN 523 561
DR DOMAIN 565 604
DR DOMAIN 615 854
DR ACT_SITE 656 656
DR ACT_SITE 711 711
DR ACT_SITE 805 805
DR CARBOHYD 107 107
DR CARBOHYD 302 302
DR CARBOHYD 365 365
DR CARBOHYD 421 421
DR CARBOHYD 489 489
DR CARBOHYD 772 772
DR SEQUENCE 855 AA; 94654 MW; 4F10E84DA2146DD5 CRC64;
Query Match 64.6%; Score 53; DB 1; Length 855;
Best Local Similarity 62.5%; Pred. No. 0.15;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
1 VVGGTAAQGEFFPMV 16
||||| :|||:|
615 VVGGTNADEGWPMQV 630
DR
RESULT 6
VESOR
CTR2_VESOR STANDARD; PRT; 216 AA.
AC P00768;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin II (EC 3.4.21.1).
OS Vespa orientalis (Oriental hornet).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
OC Aculeata; Vespoidae; Vespidae; Vespinae; Vespa.
OX NCBI_TaxID=7447;
RN [1]
RP SEQUENCE. AND DISULFIDE BONDS.
RX MEDLINE=83178101; PubMed=6340663;
RA Jany K.-D., Bekelar K., Pfeleiderer G., Ishay J.;
RT "Amino acid sequence of an insect chymotrypsin from the larvae of the
RN hornet, Vespa orientalis."
RL Biochem. Biophys. Res. Commun. 110:1-7(1983).
[2]
RP ACTIVE SITE.
RX MEDLINE=81208311; PubMed=6786354;
RA Jany K.-D., Bekelar K., Ishay J.;
RT "The amino acid sequences around the reactive serine and histidine
RN residues of the chymotrypsin-like protease from the hornet, Vespa
```

```
RT orientalis.";
RL Biochim. Biophys. Acta 668:197-200(1981).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC Phe-|-Xaa, Leu-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR: A00954; KYH20.
DR HSSP: P00763; LDPO.
DR MEROPS: S01.121; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp SPC; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
KW ACT_SITE 39 39 CHARGE RELAY SYSTEM.
FT ACT_SITE 82 82 CHARGE RELAY SYSTEM.
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM.
FT DISULFID 25 40
FT DISULFID 146 159
FT DISULFID 169 193
FT SEQUENCE 216 AA; 23471 MW; F235BF992AEFDE1 CRC64;
Query Match 63.4%; Score 52; DB 1; Length 216;
Best Local Similarity 50.0%; Pred. No. 0.052;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGGTAAQGEFFPMV 16
||||| :|||:|
DB 1 IVGGTNAPRGKPYQV 16
DR
RESULT 7
TRYP_ASTFL STANDARD; PRT; 237 AA.
AC P00765;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trypsin I (EC 3.4.21.4).
OS Astacus fluviatilis (Broad-fingered crayfish) (Astacus astacus).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Astacidea; Astacoidea; Astacidae; Astacus.
OX NCBI_TaxID=6715;
RN [1]
RP SEQUENCE.
RX MEDLINE=83178967; PubMed=6838862;
RA Titani K., Sasaigawa T., Woodbury R.G., Ericsson L.H., Dorsam H.,
RA Kraemer M., Neurath H., Zwilling R.;
RT "Amino acid sequence of crayfish (Astacus fluviatilis) trypsin IF.";
RL Biochemistry 22:1459-1465(1983).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- MISCELLANEOUS: TRYPSIN I IS ONE OF FIVE FORMS OF THE ENZYME KNOWN
CC TO BE PRESENT IN CRAYFISH. THIS PROTEIN IS MORE ACIDIC THAN
CC MAMMALIAN TRYPSIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR: A00951; TRCY1.
DR HSSP: P00763; LDPO.
DR MEROPS: S01.122; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp SPC; 1.
DR PROSITE: PS02040; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas.
```

```

FT ACT_SITE 45 45 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 189 189 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 30 46
FT DISULFID 159 174
FT DISULFID 185 213
FT SITE 183 183
SQ SEQUENCE 237 AA; 25021 MW; 4072133E55022C76 CRC64;

Query Match 63.4%; Score 52; DB 1; Length 237;
Best Local Similarity 56.2%; Pred. No. 0.057;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVGTRAAOGEFFPMV 16
:|||||:
Db 1 IVGTDVGLGEFFYQL 16

RESULT 8
CTRL_HALRU STANDARD; PRT; 254 AA.
AC F35003;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
Chymotrypsin-like serine proteinase precursor (EC 3.4.21.-).
Haliotis rufescens (California red abalone).
Eukaryota; Metazoa; Mollusca; Gastropoda; Archaeogastropoda;
OC Haliotidae; Haliotis.
OX NCBI_TaxID=6454;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Intestine;
RX MEDLINE=93343624; PubMed=824947;
RA Groppe J.C., Morse D.E.;
RT "Molluscan chymotrypsin-like protease: structure, localization, and
substrate specificity.";
RL Arch. Biochem. Biophys. 305:159-169(1993).
CC -!- FUNCTION: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -!- ENZYME REGULATION: ACTIVATED BY AN AUTOCATALYTIC MECHANISM.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: EXTRACELLULAR.
CC -!- SUBCELLULAR LOCATION: OF THE INTESTINE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; X71438; CAA50572.1; -.
PIR; S35585; S35585.
DR HSP; P00763; LDPO.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser. protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal; Zymogen.
FT SIGNAL 1 18
FT PROPEP 19 23
FT CHAIN 24 254
FT ACT_SITE 68 68
FT ACT_SITE 117 117
CHARGE RELAY SYSTEM (BY SIMILARITY).

```

```

FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 53 69 BY SIMILARITY.
FT DISULFID 146 218 BY SIMILARITY.
FT DISULFID 181 199 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
SQ SEQUENCE 254 AA; 27250 MW; ADAA9A9A22BEFCEC CRC64;

Query Match 63.4%; Score 52; DB 1; Length 254;
Best Local Similarity 64.3%; Pred. No. 0.062;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAOGEFFPW 14
:|||||:
Db 24 IVGGSNAAGFEFW 37

RESULT 9
FA9_SHEEP STANDARD; PRT; 274 AA.
ID FA9_SHEEP
AC P16291;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
Coagulation factor IX (EC 3.4.21.22) (Christmas factor) (Fragment).
GN F9.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90152675; PubMed=2303254;
RA Sarkar G., Koeberl D.D., Sommer S.S.;
RT "Direct sequencing of the activation peptide and the catalytic domain
of the factor IX gene in six species.";
RL Genomics 6:133-143(1990).
CC -!- FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT
PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY
CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++
IONS, PHOSPHOLIPIDS, AND FACTOR VIIIa.
CC -!- CATALYTIC ACTIVITY: Hydrolyzes one Arg-|-Ile bond in factor X to
form factor Xa.
CC -!- SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIA, WHICH EXCISES THE
ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2
CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; M26233; AAA31520.1; -.
DR HSP; P16293; IPEX.
DR MEROPS; S01.214; -.
DR InterPro; IPR001254; Ser. protease_Try.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp Spc; 1.
DR PROSITE; PS00011; GUU CARBOXYLATION; PARTIAL.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Blood coagulation; Plasma; Serine protease; Calcium-binding;
Hydrolase; Glycoprotein.
FT NON_TER 1
FT ACT_SITE 89 89 CHARGE RELAY SYSTEM.
FT ACT_SITE 137 137 CHARGE RELAY SYSTEM.
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM.

```

FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 274 274
SQ SEQUENCE 274 AA; 30595 MW; D3617FC3BD33E9B CRC64;

Query Match 63.4%; Score 52; DB 1; Length 274;
Best Local Similarity 62.5%; Pred. No. 0.067; 3; Mismatches 0; Gaps 0;
Matches 10; Conservative 3; Indels 3; Gaps 0;

QY 1 VVGTTAAQGEFFPMV 16
49 VVGEDAAQGEFFPMV 64
RESULT 10
KAL MOUSE
ID KAL_MOUSE STANDARD; PRT; 638 AA.
P26262;
01-MAY-1992 (Rel. 22, Created)
01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Plasma kallikrein precursor (EC 3.4.21.34) (Plasma prekallikrein)
DE (Kininogenin) (Fletcher factor).
GN KLK1 OR KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
STRAIN=BALB/c; TISSUE=Liver;
MEDLINE=91090844; PubMed=2264928;
RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaubien G.,
RA Brachpapa L., Rochemont J., Mbikay M., Chretien M.,
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Xaa and Lys-|-Xaa
CC bonds, including Lys-|-Arg and Arg-|-Ser bonds in (human)
CC kininogen to release bradykinin.
CC -1- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. PLASMA KALLIKREIN
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL; M58588; AAA63393.1; .
CC PIR; A36557; KQMSPL.
CC HSP; P00750; 1RTF.
CC MEROPS; S01.212; .
CC MGD; MGI:102849; KIK3.
CC InterPro; IPR001177; Apple.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR003014; PAN.
CC InterPro; IPR001254; Ser_protease_Try.
CC Pfam; PF00024; PAN; 4.

DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00005; APPLEDOMAIN.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00223; APPLE; 4.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00495; APPLE; 4.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
KW Repeat.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT DOMAIN 20 105 APPLE 1.
FT DOMAIN 110 195 APPLE 2.
FT DOMAIN 200 285 APPLE 3.
FT DOMAIN 291 376 APPLE 4.
FT DOMAIN 389 621 SERINE PROTEASE.
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 453 453 O-LINKED (PROBABLE).
FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (PROBABLE).
FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
FT DISULFID 21 104 BY SIMILARITY.
FT DISULFID 47 77 BY SIMILARITY.
FT DISULFID 51 57 BY SIMILARITY.
FT DISULFID 111 194 BY SIMILARITY.
FT DISULFID 137 166 BY SIMILARITY.
FT DISULFID 141 147 BY SIMILARITY.
FT DISULFID 201 284 BY SIMILARITY.
FT DISULFID 227 256 BY SIMILARITY.
FT DISULFID 231 237 BY SIMILARITY.
FT DISULFID 292 375 BY SIMILARITY.
FT DISULFID 318 347 BY SIMILARITY.
FT DISULFID 322 328 BY SIMILARITY.
FT DISULFID 340 345 BY SIMILARITY.
FT DISULFID 383 503 BY SIMILARITY.
FT DISULFID 419 435 BY SIMILARITY.
FT DISULFID 517 584 BY SIMILARITY.
FT DISULFID 548 563 BY SIMILARITY.
FT DISULFID 574 602 BY SIMILARITY.
SQ SEQUENCE 638 AA; 71368 MW; CC27C93AB1086599 CRC64;

Query Match 62.2%; Score 51; DB 1; Length 638;
Best Local Similarity 56.2%; Pred. No. 0.24;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTTAAQGEFFPMV 16
391 IVGGTNASLGEWPMQV 406
RESULT 11
ST14_HUMAN
ID ST14_HUMAN STANDARD; PRT; 855 AA.
AC Q9Y5Y6; Q9H3S0; Q9HCA3; Q9BS01; Q9HB36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suppressor of tumorigenicity 14 (EC 3.4.21.-) (Matritase) (Membrane-
DE type serine protease 1) (Mtr-Spl) (Proctamin) (Serine protease TAG-15)
DE (Tumor associated differentially-expressed gene-15 protein).
GN ST14 OR PRSS14 OR SNC19 OR TADG15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99303581; PubMed=10373424;
 RA Lin C.Y., Anders J., Johnson M., Sang Q.A., Dickson R.B.;
 RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
 protease with trypsin-like activity";
 RL J. Biol. Chem. 274:18231-18236(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99432178; PubMed=10500122;
 RA Takeuchi T., Shuman M.A., Craik C.S.;
 RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
 dissect complex biological processes and identify a membrane-type
 serine protease in epithelial cancer and normal tissue";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Prostate;
 RA Yamaguchi N., Mitsui S.;
 RT "Molecular cloning of a novel transmembrane serine protease expressed
 in human prostate";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX Tanimoto H., Underwood L.J., Wang Y., Shigemasa K., Parmley T.H.,
 RA O'Brien T.J.;
 RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 327-855 FROM N.A.
 RX TISSUE=Muscle;
 RA Strausberg R.;
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 340-664 FROM N.A.
 RX Cao J., Fan W., Zheng S.;
 RA "Genomic analysis of a novel human serine protease SN19";
 RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP CHARACTERIZATION.
 RX TISSUE=Milk;
 RA MEDLINE=99303582; PubMed=10373425;
 RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
 RT "Purification and characterization of a complex containing matriptase
 and a Kunitz-type serine protease inhibitor from human milk";
 RL J. Biol. Chem. 274:18237-18242(1999).
 CC -!- FUNCTION: DEGRADES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
 IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
 ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
 OR LYS AS THE P1 SITE.
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF118224; RAD42765.2; -;
 DR EMBL; AF133086; AAF00109.1; -;
 DR EMBL; AF030036; BAB20376.1; -;
 DR EMBL; AF057145; AAG15395.1; -;
 DR EMBL; BC005826; AAG05826.1; -;
 DR EMBL; AF283256; AAG13949.1; -;
 DR HSSP; P00763; IDPO. -;
 DR Genem; HGNC:11344; ST14.
 DR MIM; 606797; -;
 DR MEROPS; S01.302; -;
 DR InterPro; IPR000859; CUB domain.
 DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR002172; LDL recept A.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00057; ldl_recept_a; 4.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00431; CUB; 2.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00261; LDLRECEPTOR.
 DR SMART; SM00042; CUB; 2.
 DR SMART; SM00192; LDLa; 3.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Signal-anchor; Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55
 FT TRANSMEM 56 76
 FT CYTOPLASMIC (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT EXTRACELLULAR (POTENTIAL).
 FT CUB 1.
 FT CUB 2.
 FT LDL-RECEPTOR CLASS A 1.
 FT LDL-RECEPTOR CLASS A 2.
 FT LDL-RECEPTOR CLASS A 3.
 FT LDL-RECEPTOR CLASS A 4.
 FT SERINE PROTEASE.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT FEA -> GTR (IN REF. 5).
 FT R -> S (IN REF. 4).
 FT A -> V (IN REF. 3).
 FT CONFLICT 381 381
 FT CONFLICT 674 674
 SQ SEQUENCE 855 AA; 94769 MW; 26143132C01F99C9 CRC64;
 Query Match 62.2%; Score 51; DB 1; Length 855;
 Best Local Similarity 62.5%; Pred. No. 0.33;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 VVGGTAAQGEFFPMV 16
 Db 615 VVGGTDADEGEWPMQV 630
 RESULT 12
 TRYP_PENMO
 ID TRYP_PENMO STANDARD; PRT; 32 AA.
 AC P35050;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Trypsin (EC 3.4.21.4) (Fragment).
 OS Penaeus monodon (Penaeid shrimp).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
 OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata;
 OC Penaeoidea; Penaeidae; Penaeus.
 OC NCBI_TaxID=6687;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Midgut;
 RX MEDLINE=91151586; PubMed=1963309;
 RA Lu P.J., Liu H.C., Tsai I.H.;
 RT "The midgut trypsins of shrimp (Penaeus monodon). High efficiency
 toward native protein substrates including collagens";
 RL Biol. Chem. Hoppe-Seyler 371:851-859(1990).
 CC -!- FUNCTION: HIGH EFFICIENCY TOWARD NATIVE PROTEIN SUBSTRATES
 CC INCLUDING COLLAGENS. CC

```
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
DR PIR; S11537; S11537.
DR MEROPS; S01.111; -.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; PARTIAL.
DR PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
DR PROSITE; PS00135; TRYPSIN_SER; PARTIAL.
KW Hydrolase; Serine protease.
FT NON_TER 32
FT SEQUENCE 32 AA; 3330 MW; 96A3A5D1F4B19182 CRC64;
Query Match 61.0%; Score 50; DB 1; Length 32;
Best Local Similarity 50.0%; Pred. No. 0.016;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 VVGTRAAQGEFFPMV 16
1 IVGTAVTPGEFFYQL 16
RESULT 13
ID_CTR2_VESCR STANDARD; PRT; 218 AA.
AC P00769; 1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chymotrypsin II (EC 3.4.21.1).
OS Vespa crabro (European hornet).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apoecrita;
OC Aculeata; Vespoidea; Vespidae; Vespinae; Vespa.
ON NCBI_TaxID=7445;
RN 1
RP SEQUENCE.
RA Jany K.-D., Haug H.;
RT "Amino acid sequence of the chymotryptic protease II from the larvae
RT of the hornet, Vespa crabro.";
RT FEBS Lett. 158:98-102(1983).
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-|-Xaa, Trp-|-Xaa,
CC phe-|-Xaa, Leu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: AN ADDITIONAL ARG AT THE CARBOXYL END WAS FOUND IN
CC SOME OF THE MOLECULES.
CC PIR; A00955; KYVH2C.
CC HSP; P00763; 1DPO.
CC MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT ACT_SITE 39 CHARGE RELAY SYSTEM.
FT ACT_SITE 84 CHARGE RELAY SYSTEM.
FT ACT_SITE 175 CHARGE RELAY SYSTEM.
FT DISULFID 25 40
FT DISULFID 148 161
FT DISULFID 171 195
FT SEQUENCE 218 AA; 23677 MW; 509AB50DE190EB39 CRC64;
Query Match 61.0%; Score 50; DB 1; Length 218;
Best Local Similarity 50.0%; Pred. No. 0.12;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGTRAAQGEFFPMV 16
```

```
Db 1 IVGGTDAPRGKYPQV 16
RESULT 14
CAP7_PIG STANDARD; PRT; 219 AA.
ID CAP7_PIG
AC P80015;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Azurocidin (Cationic antimicrobial protein CAP37) (Heparin-binding
DE protein) (HBP).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN 1
RP SEQUENCE.
RC TISSUE=Neutrophils;
RX MEDLINE=91224149; PubMed=2026172;
RA Flodgaard H., Oestergaard E., Bayne S., Svendsen A., Thomsen J.,
RA Engels M., Wollmer A.;
RT "Covalent structure of two novel neutrophil leucocyte-derived
RT proteins of porcine and human origin. Neutrophil elastase homologues
RT with strong monocyte and fibroblast chemotactic activities.";
RL Eur. J. Biochem. 197;535-547(1991).
CC -1- FUNCTION: THIS IS A NEUTROPHIL GRANULE-DERIVED ANTIBACTERIAL AND
CC MONOCYTE- AND FIBROBLAST-SPECIFIC CHEMOTACTIC GLYCOPROTEIN. BINDS
CC HEPARIN.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ELASTASE SUBFAMILY.
DR PIR; S15393; TRPGAZ.
DR HSP; P20160; 1A7S.
DR MEROPS; S01.971; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
KW Serine protease homology; Glycoprotein; Chemotaxis; Antibiotic;
KW Heparin-binding.
FT CARBOHYD 113 113 N-LINKED (GLCNAC. .).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .).
FT DISULFID 26 42 BY SIMILARITY.
FT DISULFID 122 179 BY SIMILARITY.
FT DISULFID 152 158 BY SIMILARITY.
FT SEQUENCE 219 AA; 24301 MW; 99129CAD88B0749D CRC64;
Query Match 59.8%; Score 49; DB 1; Length 219;
Best Local Similarity 60.0%; Pred. No. 0.18;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 VVGTRAAQGEFFPMV 15
1 IVGRRRAQGEFFFL 15
RESULT 15
ACRO_RABIT STANDARD; PRT; 431 AA.
ID ACRO_RABIT
AC P48038;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Acrosin precursor (EC 3.4.21.10).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
```



```

OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Testis;
RX MEDLINE=94368861; PubMed=8086468;
RA Richardson R.T., O'Rand M.G.;
RT "Cloning and sequencing of cDNAs for rabbit preproacrosin and a novel
RL preproacrosin-related cDNA.";
RL Biochim. Biophys. Acta 1219:215-218(1994).
CC -!- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.
CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY. IT
CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
CC ACROSOME.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL
CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.
CC -!- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U05204; AAA61630.1; -.
DR HSP; P00763; LDPO.
DR MEROPS; S01.223; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYD_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal.
FT SIGNAL 1 16 BY SIMILARITY.
FT CHAIN 17 431 ACROSIN.
FT CHAIN 17 39 ACROSIN LIGHT CHAIN (BY SIMILARITY).
FT CHAIN 40 ? ACROSIN HEAVY CHAIN (BY SIMILARITY).
FT PROPEP ? 431 PRO-RICH.
FT DOMAIN 40 288 SERINE PROTEASE.
FT DISULFID 22 152 INTERCHAIN (BY SIMILARITY).
FT DISULFID 26 160 INTERCHAIN (BY SIMILARITY).
FT DISULFID 71 87 BY SIMILARITY.
FT DISULFID 175 244 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 234 264 BY SIMILARITY.
FT CARBOHYD 19 19 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 431 AA; 46422 MW; 1C015A4E0BC0C668 CRC64;

Query Match 59.8%; Score 49; DB 1; Length 431;
Best Local Similarity 62.5%; Pred. No. 0.37;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VVGGTAAQGGFFPMV 16
Db 40 VVGQAQQAQGAFFMV 55

```

Search completed: April 11, 2003, 17:57:30
Job time : 12 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:56:10 ; Search time 16 Seconds
(without alignments)
96.134 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVGGTAAQGEFFFMV 16
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 283224 seqs, 96134422 residues
1 number of hits satisfying chosen parameters: 283224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	259	1 TRSMG	trypsin (EC 3.4.21
2	60.5	73.8	18	2 A61577	24k serine protein
3	60	73.2	254	2 S49329	trypsin-like prote
4	55	67.1	281	2 T13596	trypsin homolog -
5	54	65.9	24	2 PN0655	alkaline trypsin-1
6	53	64.6	855	2 JC7731	membrane-bound arg
7	53	64.6	855	2 JC7775	membrane type-seri
8	52	63.4	20	2 A61327	trypsin (EC 3.4.21
9	52	63.4	216	1 KYVH20	chymotrypsin (EC 3
10	52	63.4	237	1 TRCV1	trypsin (EC 3.4.21
11	52	63.4	254	2 S35885	chymotrypsin-like
12	52	63.4	274	2 I47078	coagulation factor
13	52	63.4	548	2 D82175	coagulable trypsin V
14	51	62.2	638	1 KQMSPL	plasma kallikrein
15	50	61.0	32	2 S11537	trypsin (EC 3.4.21
16	50	61.0	218	1 KYVH2C	chymotrypsin (EC 3
17	49	59.8	24	2 PN0653	alkaline trypsin-1
18	49	59.8	219	1 TRFGAZ	azurocidin - pig
19	49	59.8	431	2 S47538	acrosin (EC 3.4.21
20	49	59.8	613	2 S15468	complement C3b/C4b
21	49	59.8	625	1 KFHU1	coagulation factor
22	48	58.5	20	2 D34817	collagenolytic pro
23	48	58.5	638	1 KQRTPL	plasma kallikrein
24	48	58.5	638	1 KQRTPL	plasma kallikrein
25	47	57.3	37	2 S02176	acrosin (EC 3.4.21
26	47	57.3	266	2 S54146	trypsin (EC 3.4.21
27	47	57.3	416	1 KFB0	coagulation factor
28	47	57.3	819	2 B87580	conserved hypothet
29	46	56.1	271	2 I46580	factor IX - pig (f

30	46	56.1	275	2 I46712	factor IX - rabbit
31	46	56.1	705	1 C1HURB	complement subcomp
32	45	54.9	20	2 B34817	collagenolytic pro
33	45	54.9	20	2 C34817	collagenolytic pro
34	45	54.9	42	2 B61143	trypsin (EC 3.4.21
35	45	54.9	189	2 F69473	signal sequence pe
36	45	54.9	267	1 ELHUL	leukocyte elastase
37	45	54.9	285	2 I48144	coagulation factor
38	45	54.9	452	1 A30351	coagulation factor
39	45	54.9	461	1 KFHU	coagulation factor
40	44	53.7	583	2 A29154	complement factor
41	44	53.7	1019	1 A56318	enteropeptidase (E
42	43	52.4	250	2 S55493	serine proteinase
43	43	52.4	251	1 TRHUAZ	azurocidin precurs
44	43	52.4	415	1 A34170	acrosin (EC 3.4.21
45	43	52.4	421	1 S11674	acrosin (EC 3.4.21

ALIGNMENTS

RESULT 1

TRSMG

trypsin (EC 3.4.21.4) precursor - Streptomyces griseus
C:Species: Streptomyces griseus
C>Date: 24-Apr-1984 #sequence_revision 12-May-1994 #text_change 07-May-1999
C:Accession: JQ1302; A00962
R:Kim, J.C.; Cha, S.H.; Jeong, S.T.; Oh, S.K.; Byun, S.M.
Biochem. Biophys. Res. Commun. 181, 707-713, 1991
A:Title: Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin gene
A:Reference number: JQ1302; MUID:92095977; PMID:1755852
A:Accession: JQ1302
A:Molecule type: DNA
A:Residues: 1-259 <KIM>
A:Cross-references: GB:M64471
R:Olafson, R.W.; Jurasek, L.; Carpenter, M.R.; Smillie, L.B.
Biochemistry 14, 1168-1177, 1975
A:Title: Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragment
A:Reference number: A00962; MUID:75127940; PMID:804314
A:Accession: A00962
A:Molecule type: protein
A:Residues: 37-95, 98-259 <OLA>
R:Read, R.J.; James, M.N.G.
J. Mol. Biol. 200, 523, 1988
A:Title: Refined crystal structure of Streptomyces griseus trypsin at 1.7 angstroms res
A:Reference number: A44574; MUID:88286735; PMID:3135412
A:Contents: annotation; X-ray crystallography, 1.7 angstroms
A:Note: residues 96-97 modeled as Gly-Ala
C:Genetics:
A:Gene: sprT
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F1-32/Domain: signal sequence #status predicted <SIG>
F33-36/Domain: propeptide #status predicted <PRO>
F37-258/Product: trypsin #status experimental <MAT>
F37-252/Domain: trypsin homology <TRY>
F58-74,177-192,204-233/disulfide bonds: #status experimental
F73,118,208/Active site: His, Asp, Ser #status experimental

Query Match 100.0%; Score 82; DB 1; Length 259;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFFMV 16
|||
Db 37 VVGGTAAQGEFFFMV 52

RESULT 2

A61577

24k serine proteinase (EC 3.4.21.-) - Streptomyces fradiae (fragment)
C:Species: Streptomyces fradiae

C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 06-Jan-1995

C;Accession: A61577

R;Sinha, U.; Wolz, S.A.; Lad, P.J.

Int. J. Biochem. 23, 979-994, 1991

A;Title: Two new extracellular serine proteases from *Streptomyces fradiae*.

A;Reference number: A61577; MUID:92155439; PMID:1786859

C;Accession: A61577

A;Molecule type: protein

A;Residues: 1-18 <SIN>

C;Keywords: extracellular protein; hydrolase; serine proteinase

Query Match 73.8%; Score 60.5; DB 2; Length 18;

Best Local Similarity 87.5%; Pred. No. 0.0006;

Matches 14; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 VVGGTAAAGGEPFPMV 16

||||| |||||

Db 1 VVGGTAAAGGEPFPMV 15

ULT 3

S49329

trypsin-like proteinase 1 (EC 3.4.21.-) precursor - imperfect fungus (*Metarhizium anisopliae*)

C;Species: *Metarhizium anisopliae*

C;Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 22-Jun-1999

C;Accession: JC4517; S49329

R;Smithson, S.L.; Paterson, I.C.; Bailey, A.M.; Screen, S.E.; Hunt, B.A.; Cobb, B.D.; Co

Gene 166, 161-169, 1995

A;Title: Cloning and characterisation of a gene encoding a cuticle-degrading protease fr

A;Reference number: JC4517; MUID:96105219; PMID:8529882

C;Accession: JC4517

A;Molecule type: DNA

A;Residues: 1-254 <SM2>

A;Cross-references: EMBL:X78875; NID:G556656; PIDN:CAA55477.1; PID:G556657

A;Experimental source: ME1

A;Note: submitted to the EMBL Data Library, April 1994

C;Genetics:

A;Gene: trv1

A;Introns: 86/3; 149/2

C;Superfamily: trypsin; trypsin homology

C;Keywords: endoplasmic reticulum; hydrolase; serine proteinase

21/Domain: signal sequence #status predicted <SIG>

29/Domain: propeptide #status predicted <PRP>

30-254/Product: trypsin-like proteinase #status predicted <MAT>

P;30-250/Domain: trypsin homology <TRY>

P;70,112,209/Active site: His, Asp, Ser #status: predicted

Query Match 73.2%; Score 60; DB 2; Length 254;

Best Local Similarity 68.8%; Pred. No. 0.01;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16

||||| |||||

Db 30 IVGGSPAAAGGEPFIV 45

RESULT 4

TI13596

trypsin homolog - fruit fly (*Drosophila melanogaster*)

N;Alternate names: protein EG:80H7.1

C;Species: *Drosophila melanogaster*

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C;Accession: TI13596

R;Benos, P.

submitted to the EMBL Data Library, April 1999

A;Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.

A;Reference number: Z17667

C;Accession: TI13596

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-281 <BEN>

A;Cross-references: EMBL:AL031027; PIDN:CAA19843.2

C;Genetics:

A;Cross-references: FlyBase:FBgn0000481

A;Introns: 159/3

A;Note: EG:80H7.1

C;Superfamily: trypsin; trypsin homology

Query Match 67.1%; Score 55; DB 2; Length 281;

Best Local Similarity 62.5%; Pred. No. 0.082;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16

||||| |||||

Db 27 IVNGTTAGGEPFPMV 42

RESULT 5

PN0655

alkaline trypsin-like serine proteinase (EC 3.4.21.-) F-II - earthworm (*Lumbricus rubell*

N;Alternate names: fibrinolytic enzyme

C;Species: *Lumbricus rubellus* (humus earthworm)

C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 07-May-1999

C;Accession: PN0655

R;Nakajima, N.; Mihara, H.; Sumi, H.

Biosci. Biotechnol. Biochem. 57, 1726-1730, 1993

A;Title: Characterization of potent fibrinolytic enzymes in earthworm, *Lumbricus rubell*,

A;Reference number: PN0653; MUID:94080032; PMID:7764268

C;Accession: PN0655

A;Molecule type: protein

A;Residues: 1-24 <NAK>

C;Comment: This enzyme is a stable and potent fibrinolytic enzyme and is absorbed from

C;Keywords: hydrolase; serine proteinase

Query Match 65.9%; Score 54; DB 2; Length 24;

Best Local Similarity 56.2%; Pred. No. 0.01;

Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16

||||| |||||

Db 1 VVGGTAAAGGEPFPMV 16

RESULT 6

JC7731

membrane-bound arginine-specific serine proteinase precursor - rat

C;Species: *Rattus norvegicus* (Norway rat)

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 14-Dec-2001

C;Accession: JC7731

R;Kishi, K.; Yamazaki, K.; Yasuda, I.; Yahagi, N.; Ichinose, M.; Tsuchiya, Y.; Athauda,

J. Biochem. 130, 425-430, 2001

A;Title: Characterization of a membrane-bound arginine-specific serine proteinase from ra

A;Reference number: JC7731; MUID:21421307; PMID:11530019

C;Accession: JC7731

A;Molecule type: mRNA

A;Residues: 1-855 <KIS>

A;Cross-references: DDBJ:AB049189

A;Experimental source: strain Male, 7-week-old

C;Comment: This enzyme, localized mainly on brushborder membranes of the intestine, par

C;Keywords: protein digestion

Query Match 64.6%; Score 53; DB 2; Length 855;

Best Local Similarity 62.5%; Pred. No. 0.55;

Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAAGGEPFPMV 16

||||| |||||

Db 615 VVGGTAAAGGEPFPMV 630

RESULT 7

JC7775

membrane type-serine protease 1 - rat

C;Species: *Rattus norvegicus* (Norway rat)

C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C;Accession: JC7775

R;Satomi, S.; Yanasaki, Y.; Tsuchi, S.; Hitomi, Y.; Iwanaga, T.; Fushiki, T.
 Biochem. Biophys. Res. Commun. 287, 995-1002, 2001
 A;Title: A role for membrane-type serine protease
 A;Reference number: JC7775; PMID:11573963
 A;Contents: Small intestine
 A;Accession: JC7775
 A;Molecule type: mRNA
 A;Residues: 1-855 <SAT>
 A;Cross-references: DDBJ:AB037898
 C;Comment: This enzyme, an epithelial-derived, type II integral membrane serine protease
 ciated with epithelial migration and/or cell loss not only as an upstream activator of u
 proteins.
 C;Genetics:
 A;Gene: mt-spl
 A;Map position: basolateral cell surface

Query Match 64.6%; Score 53; DB 2; Length 855;
 Best Local Similarity 62.5%; Pred. No. 0.55; Mismatches 3; Indels 0; Gaps 0;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 615 VVGGTNADEGEWPMQV 630

RESULT 8
 TRCY1
 trypsin (EC 3.4.21.4) - narrow-fingered crayfish (fragment)
 C;Species: Astacus leptodactylus (narrow-fingered crayfish)
 C;Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
 C;Accession: A61327
 R;Zwilling, R.; Neurath, H.; Ericsson, L.H.; Enfield, D.L.
 FEBS Lett. 60, 247-249, 1975
 A;Title: The amino-terminal sequence of an invertebrate trypsin (crayfish Astacus leptodactylus)
 A;Reference number: A61327; MUID:76210806; PMID:1227963
 A;Accession: A61327
 A;Molecule type: protein
 A;Residues: 1-20 <ZWI>
 C;Comment: This enzyme could not be demonstrated as a zymogen, even in homogenates of th
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase

Query Match 63.4%; Score 52; DB 2; Length 20;
 Best Local Similarity 56.2%; Pred. No. 0.019; Mismatches 3; Indels 0; Gaps 0;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 1 IVGTDATLGEFFPYQL 16

RESULT 9
 KVVH20
 chymotrypsin (EC 3.4.21.1) II - oriental hornet
 C;Species: Vespa orientalis (oriental hornet)
 C;Date: 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change 18-Jul-1997
 C;Accession: A00954
 R;Jany, K.D.; Bekelar, K.; Pfeleiderer, G.; Ishay, J.
 Biochem. Biophys. Res. Commun. 110, 1-7, 1983
 A;Title: Amino acid sequence of an insect chymotrypsin from the larvae of the hornet, V
 A;Reference number: A90109; MUID:83178101; PMID:6340663
 A;Accession: A00954
 A;Molecule type: protein
 A;Residues: 1-216 <JAN>
 R;Jany, K.D.; Bekelar, K.; Ishay, J.
 Biochim. Biophys. Acta 668, 197-200, 1981
 A;Title: The amino acid sequences around the reactive serine and histidine residues of b
 A;Reference number: A90635; MUID:81208311; PMID:6786354
 A;Contents: annotation; active site
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase
 F;1-211/Domain: trypsin homology <TRY>
 F;25-40,146-159,169-193/Disulfide bonds: #status experimental

F;39,173/Active site: His, Ser #status experimental
 F;82/Active site: Asp #status predicted

Query Match 63.4%; Score 52; DB 1; Length 216;
 Best Local Similarity 50.0%; Pred. No. 0.21; Mismatches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 1 IVGTTNAPRGKYPYQV 16

RESULT 10
 TRCY1
 trypsin (EC 3.4.21.4) I - broad-fingered crayfish
 C;Species: Astacus astacus, Astacus fluviatilis (broad-fingered crayfish)
 C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 18-Jul-1997
 C;Accession: A00951
 R;Titani, K.; Sasagawa, T.; Woodbury, R.G.; Ericsson, L.H.; Dorsam, H.; Kraemer, M.; Nei
 Biochemistry 22, 1459-1465, 1983
 A;Title: Amino acid sequence of crayfish Astacus fluviatilis trypsin I-f.
 A;Reference number: A00951; MUID:83178967; PMID:6838862
 A;Accession: A00951
 A;Molecule type: protein
 A;Residues: 1-237 <TIT>
 C;Comment: Trypsin I is one of five forms of the enzyme known to be present in crayfish
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; protein digestion; serine proteinase
 F;1-232/Domain: trypsin homology <TRY>
 F;30-46,159-174,185-213/Disulfide bonds: #status predicted
 F;45,96,189/Active site: His, Asp, Ser #status predicted

Query Match 63.4%; Score 52; DB 1; Length 237;
 Best Local Similarity 56.2%; Pred. No. 0.23; Mismatches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 1 IVGTTDAVLGEFFPYQL 16

RESULT 11
 S35585
 chymotrypsin-like proteinase (EC 3.4.21.-) precursor - California red abalone
 C;Species: Haliotis rufescens (California red abalone)
 C;Date: 07-Apr-1994 #sequence_revision 19-May-1994 #text_change 22-Jun-1999
 C;Accession: S35585; S32750
 R;Groppe, J.C.; Morse, D.E.
 Arch. Biochem. Biophys. 305, 159-169, 1993
 A;Title: Molluscan chymotrypsin-like protease: structure, localization, and substrate s
 A;Reference number: S35585; MUID:93343624; PMID:8342947
 A;Accession: S35585
 A;Molecule type: mRNA
 A;Residues: 1-254 <GRO>
 A;Cross-references: EMBL:X71438; NID:G296361; PIDN:CAA50572.1; PID:G296362
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: hydrolase; intestine; serine proteinase; zymogen
 F;1-23/Domain: signal sequence #status predicted <SIG>
 F;23-254/Product: chymotrypsin-like proteinase #status predicted <MAT>
 F;24-251/Domain: trypsin homology <TRY>

Query Match 63.4%; Score 52; DB 2; Length 254;
 Best Local Similarity 64.3%; Pred. No. 0.24; Mismatches 3; Indels 0; Gaps 0;

1 VVGTRAAQGEFFPMV 14
 :|||:|:|:|:
 24 IVGGNNAAAGEFFPMV 37

RESULT 12
 147078
 coagulation factor IXa (EC 3.4.21.22) - sheep (fragment)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:57:21 ; Search time 14 Seconds
(without alignments)
69.870 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVGCTRAAQGEFFPMV 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 248812 seqs, 61136040 residues

1 number of hits satisfying chosen parameters: 248812

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	64.6	855	10	US-09-900-751-2
2	52	63.4	20	9	US-09-938-269-13
3	51	62.2	241	9	US-10-099-700A-2
4	51	62.2	241	9	US-10-092-004A-2
5	51	62.2	620	10	US-09-925-301-1193
6	51	62.2	855	9	US-10-099-700A-2
7	50	61.0	25	9	US-09-938-269-4
8	49	59.8	625	9	US-09-808-602-104
9	48	58.5	638	9	US-09-938-269-12
10	48	58.5	1128	9	US-09-808-602-102
11	48	58.5	1128	10	US-09-888-615-97
12	46	56.1	449	10	US-09-925-302-612
13	46	56.1	688	10	US-09-874-198-7
14	46	56.1	688	10	US-09-874-238-7
15	45	54.9	18	10	US-09-808-602-94
16	45	54.9	18	10	US-09-861-708-2
17	45	54.9	20	9	US-09-938-269-10
18	45	54.9	20	9	US-09-938-269-11
19	45	54.9	218	10	US-09-861-708-3

20 45 54.9 246 9 US-09-854-133-60 Sequence 60, Appl
21 45 54.9 246 10 US-09-738-973-60 Sequence 60, Appl
22 45 54.9 305 9 US-10-067-761-12 Sequence 12, Appl
23 45 54.9 305 10 US-09-804-156-12 Sequence 12, Appl
24 45 54.9 407 12 US-10-060-333-4 Sequence 4, Appl
25 45 54.9 415 10 US-09-118-748-2 Sequence 2, Appl
26 45 54.9 418 9 US-09-854-133-62 Sequence 62, Appl
27 45 54.9 418 9 US-09-854-133-82 Sequence 82, Appl
28 45 54.9 418 9 US-09-854-133-83 Sequence 83, Appl
29 45 54.9 418 10 US-09-738-973-82 Sequence 62, Appl
30 45 54.9 418 10 US-09-738-973-82 Sequence 82, Appl
31 45 54.9 418 10 US-09-738-973-83 Sequence 82, Appl
32 45 54.9 423 9 US-10-063-547-106 Sequence 106, Appl
33 45 54.9 423 9 US-10-174-590-320 Sequence 320, App
34 45 54.9 423 9 US-10-176-758-320 Sequence 320, App
35 45 54.9 423 9 US-10-063-616-106 Sequence 106, App
36 45 54.9 423 9 US-10-175-737-320 Sequence 320, App
37 45 54.9 423 9 US-10-063-502-106 Sequence 106, App
38 45 54.9 423 9 US-10-173-706-320 Sequence 320, App
39 45 54.9 423 9 US-10-175-738-320 Sequence 320, App
40 45 54.9 423 9 US-10-175-752-320 Sequence 320, App
41 45 54.9 423 9 US-10-176-482-320 Sequence 320, App
42 45 54.9 423 9 US-10-176-757-320 Sequence 320, App
43 45 54.9 423 9 US-10-176-913-320 Sequence 320, App
44 45 54.9 423 9 US-10-180-552-320 Sequence 320, App
45 45 54.9 423 9 US-10-180-557-320 Sequence 320, App

ALIGNMENTS

RESULT 1
US-09-900-751-2
; Sequence 2, Application US/09900751
; Patent No. US20020026653A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D
; APPLICANT: Leviten, Michael W.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING SERINE
; TITLE OF INVENTION: PROTEASE GENE DISRUPTIONS
; FILE REFERENCE: R-386
; CURRENT APPLICATION NUMBER: US/09/900,751
; CURRENT FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: US 60/217,449
; PRIOR FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: US 60/223,170
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 60/223,460
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-900-751-2
Query Match 64.6%; Score 53; DB 10; Length 855;
Best Local Similarity 62.5%; Pred. No. 0.67;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 VVGCTRAAQGEFFPMV 16
Db 615 VVGCTNADEGEWPCV 630
RESULT 2
US-09-938-269-13
; Sequence 13, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.

```

; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Crayfish protease
US-09-938-269-13

Query Match      63.4%; Score 52; DB 9; Length 20;
Best Local Similarity 56.2%; Pred. No. 0.019;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 IVGTDATLGEFFIQL 16

RESULT 3
US-10-099-700A-4
; Sequence 4, Application US/10099700A
; Publication No. US2003008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-099-700A-4

Query Match      62.2%; Score 51; DB 9; Length 241;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTDADGEWPMQV 16

RESULT 4
US-10-092-004A-2
; Sequence 2, Application US/10092004A
; Publication No. US20030050251A1
; GENERAL INFORMATION:
; APPLICANT: Corvas International, Inc.
; APPLICANT: Madison, Edwin L.
; APPLICANT: Semple, Joseph Edward
; APPLICANT: Coombs, Gary Samuel
; APPLICANT: Reiner, John Eugene
; APPLICANT: Ong, Edgar O.
; APPLICANT: Araldi, Gian Luca
; TITLE OF INVENTION: Inhibitors of Serine Protease Activity of Matritase or MTSP1
; FILE REFERENCE: 018813/0282105
; CURRENT APPLICATION NUMBER: US/10/092,004A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: PCT/US01/28137
; PRIOR FILING DATE: 2001-09-07
; CURRENT APPLICATION NUMBER: 09/657,986

```

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-092-004A-2

Query Match      62.2%; Score 51; DB 9; Length 241;
Best Local Similarity 62.5%; Pred. No. 0.38;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTDADGEWPMQV 16

RESULT 5
US-09-925-301-1193
; Sequence 1193, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1193
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (375)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (501)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (532)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (546)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1193

Query Match      62.2%; Score 51; DB 10; Length 620;
Best Local Similarity 62.5%; Pred. No. 1;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
Db 475 VVGTDADGEWPMQV 490

RESULT 6
US-10-099-700A-2
; Sequence 2, Application US/10099700A
; Publication No. US2003008372A1
; GENERAL INFORMATION:
; APPLICANT: Edwin L. Madison
; APPLICANT: Edgar O. Ong
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING TRANSMEMBRANE SERINE PROTEASE 7,
; FILE REFERENCE: 24745-1613
; CURRENT APPLICATION NUMBER: US/10/099,700A

```


; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: 60/275,592
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-099-700A-2

Query Match 62.2%; Score 51; DB 9; Length 855;
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
:||||:|:|:|:
Db 615 VVGGTDADEGEWPMQV 630

US-09-938-269-4
; Sequence 4, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Panaeus monodon tryptic
US-09-938-269-4

Query Match 61.0%; Score 50; DB 9; Length 25;
Best Local Similarity 50.0%; Pred. No. 0.051;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
:||||:|:|:|:
Db 1 VVGGTAVTGEFFPYQL 16

US-09-808-602-104
; Sequence 104, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 104
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-808-602-104

Query Match 59.8%; Score 49; DB 9; Length 625;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16
:||||:|:|:|:
Db 388 IVGGTASVRGEWPMQV 403

US-09-938-269-12
; Sequence 12, Application US/09938269
; Publication No. US20030007951A1
; GENERAL INFORMATION:
; APPLICANT: Franklin, Richard L.
; APPLICANT: Cowling, Didier S.P.
; APPLICANT: Hubbel, Jeffrey A.
; APPLICANT: van de Wetering, Petra
; TITLE OF INVENTION: Treatment of Trauma
; FILE REFERENCE: 314572-105
; CURRENT APPLICATION NUMBER: US/09/938,269
; CURRENT FILING DATE: 2001-08-23
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Kamchatka crab
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(20)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-938-269-12

Query Match 58.5%; Score 48; DB 9; Length 20;
Best Local Similarity 50.0%; Pred. No. 0.087;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFP 14
:||||:|:|:|:
Db 1 IVGSGEATSGQEPY 14

US-09-808-602-102
; Sequence 102, Application US/09808602
; Patent No. US20020155115A1
; GENERAL INFORMATION:
; APPLICANT: Vernet, Corine A
; APPLICANT: Fernandes, Elma
; APPLICANT: Shimkets, Richard A
; APPLICANT: Herrman, John L
; APPLICANT: Majumder, Kumud
; APPLICANT: Mishra, Vishnu
; APPLICANT: Mezes, Peter S
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: No. US20020155115A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 15966-697 CIP
; CURRENT APPLICATION NUMBER: US/09/808,602
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/800,198
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/186,596
; PRIOR FILING DATE: 2000-03-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 102

Qy 1 VVGTRAAQGEFFPMV 16
 :||:|:|:|
 Db 447 IIGQKAKMGNFPMQV 462

RESULT 15

US-09-808-602-94
 ; Sequence 94, Application US/09808602
 ; Patent No. US20020155115A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Vernet, Corine A
 ; APPLICANT: Fernandes, Elma
 ; APPLICANT: Shimkets, Richard A
 ; APPLICANT: Herriman, John L
 ; APPLICANT: Majumder, Kumud
 ; APPLICANT: Mishra, Vishnu
 ; APPLICANT: Mezes, Peter S
 ; APPLICANT: MacDougall, John
 ; TITLE OF INVENTION: No. US20020155115A1e1 Proteins and Nuclec Acids Encoding Same
 ; FILE REFERENCE: 15966-697 CIP
 ; CURRENT APPLICATION NUMBER: US/09/808,602
 ; CURRENT FILING DATE: 2001-03-14
 ; PRIOR APPLICATION NUMBER: 09/800,198
 ; PRIOR FILING DATE: 2001-03-05
 ; PRIOR APPLICATION NUMBER: 60/186,596
 ; PRIOR FILING DATE: 2000-03-03
 ; NUMBER OF SEQ ID NOS: 114
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 94
 ; LENGTH: 705
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-808-602-94

Query Match 56.1%; Score 46; DB 9; Length 705;
 Best Local Similarity 43.8%; Pred. No. 8;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VVGTRAAQGEFFPMV 16
 :||:|:|:|
 Db 464 IIGQKAKMGNFPMQV 479

Search completed: April 11, 2003, 17:59:11
 Job time : 15 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:56:35 ; Search time 15 Seconds
(without alignments)
31.384 Million cell updates/sec

Title: US-10-006-223-1

Perfect score: 82

Sequence: 1 VVGTRAAQGEFFFW 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

1 number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	223	1	US-08-378-091-13
2	82	100.0	223	1	US-08-483-859-13
3	82	100.0	223	1	US-08-472-173-13
4	82	100.0	223	2	US-08-487-167-13
5	82	100.0	223	2	US-08-482-816-13
6	82	100.0	223	2	US-08-296-149-13
7	82	100.0	223	2	US-08-801-499-13
8	82	100.0	223	2	US-08-615-271-13
9	82	100.0	223	3	US-09-074-660-13
10	82	100.0	223	3	US-09-074-659-13
11	82	100.0	223	3	US-09-106-468-13
12	82	100.0	223	4	US-09-106-466A-13
13	82	100.0	223	4	US-09-106-467-13
14	59	72.0	224	1	US-08-553-516-2
15	59	72.0	248	1	US-08-238-130-2
16	59	72.0	248	2	US-08-921-426-4
17	59	72.0	248	3	US-08-816-915-4
18	59	72.0	248	5	PCT-US95-07743-4
19	53	64.6	902	4	US-09-644-600-10
20	52	63.4	20	2	US-08-385-540A-15
21	52	63.4	20	2	US-08-600-273A-15
22	52	63.4	20	3	US-08-486-820-15
23	52	63.4	20	4	US-09-220-731-15
24	52	63.4	237	4	US-08-163-919A-3
25	52	63.4	237	5	PCT-US94-14073-3
26	51	62.2	855	2	US-09-027-337-2
27	51	62.2	855	4	US-09-644-600-2

28	50	61.0	25	2	US-08-385-540A-5	Sequence 5, Appli
29	50	61.0	25	2	US-08-600-273A-5	Sequence 5, Appli
30	50	61.0	25	3	US-08-486-820-5	Sequence 5, Appli
31	50	61.0	25	4	US-09-220-731-5	Sequence 5, Appli
32	50	61.0	218	4	US-09-578-303-3	Sequence 3, Appli
33	49	59.8	219	2	US-08-925-708-2	Sequence 2, Appli
34	49	59.8	238	4	US-08-944-483-64	Sequence 64, Appl
35	48	58.5	20	2	US-08-385-540A-14	Sequence 14, Appl
36	48	58.5	20	3	US-08-600-273A-14	Sequence 14, Appl
37	48	58.5	20	3	US-08-486-820-14	Sequence 14, Appl
38	48	58.5	20	4	US-09-220-731-14	Sequence 14, Appl
39	48	58.5	248	4	US-08-944-483-63	Sequence 3, Appli
40	48	58.5	638	2	US-08-681-151-3	Sequence 3, Appli
41	47	57.3	23	1	US-08-485-455D-7	Sequence 7, Appli
42	47	57.3	23	2	US-08-482-130C-7	Sequence 7, Appli
43	47	57.3	23	3	US-08-484-211C-7	Sequence 7, Appli
44	47	57.3	23	3	US-08-906-769-7	Sequence 7, Appli
45	47	57.3	23	3	US-08-906-616-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-278-091-13
; Sequence 13, Application US/08278091
; Patent No. 5506139
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: COMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-278-091-13

Query Match 100.0%; Score 82; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFFW 16
|||||

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-494
TELECOMMUNICATION INFORMATION:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-494
TELECOMMUNICATION INFORMATION:

RESULT 7

US-08-801-499-13
; Sequence 13, Application US/08801499
; Patent No. 5962430
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease Act
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,499
; FILING DATE: 26-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/482,816
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-671 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1163
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; 8-801-499-13
Query Match 100.0%; Score 82; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VVGGTAAAGGEPFPMV 16
Db 1 VVGGTAAAGGEPFPMV 16
RESULT 8
US-08-615-271-13
; Sequence 13, Application US/08615271
; Patent No. 5981503
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/487,167
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,271
; FILING DATE: 20-JUN-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-580
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-615-271-13
Query Match 100.0%; Score 82; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 VVGGTAAAGGEPFPMV 16
Db 1 VVGGTAAAGGEPFPMV 16
RESULT 9
US-09-074-660-13
; Sequence 13, Application US/09074660
; Patent No. 6020183
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,660
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/487,167
; APPLICATION NUMBER: US 08/487,167
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-731 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-660-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16

RESULT 10
US-09-074-659-13
Sequence 13, Application US/09074659
Patent No. 6025342
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,659
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/487,167
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-730 MIS:jb
TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-659-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16

RESULT 11
US-09-106-468-13
Sequence 13, Application US/09106468
Patent No. 6114125
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: ANALOG OF HAEMOPHILUS HIN47 WITH REDUCED
PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,468
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/615,271
FILING DATE: 20-JUN-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-825
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-106-468-13

Query Match 100.0%; Score 82; DB 3; Length 223;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
Db 1 VVGTRAAQGEFFPMV 16


```
;
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-553-516-2

Query Match 72.0%; Score 59; DB 1; Length 224;
Best Local Similarity 62.5%; Pred. No. 0.019;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEPPFMV 16
Db 1 IVGTSASAGDFFIV 16

;
; LT 15
; US-08-238-130-2
; Sequence 2, Application US/08238130
; Patent No. 5702934
; GENERAL INFORMATION:
; APPLICANT: Hastrup, Sven
; APPLICANT: Branner, Sven
; APPLICANT: Jorgensen, Birthe R.
; APPLICANT: Christensen, Tove
; APPLICANT: Jorgensen, Birgitte B.
; APPLICANT: Shuster, Jeffrey R.
; APPLICANT: Madden, Mark
; APPLICANT: Moyer, Donna L.
; APPLICANT: Fuhsang, Claus
; TITLE OF INVENTION: PROCESSES FOR PRODUCING AN ENZYME
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5702934o No. 5702934disk of No. 5702934th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/238,130
; FILING DATE: 04-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 522/93
; FILING DATE: 05-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis Dr., Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3965.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 25..248
; FEATURE:
;
; NAME/KEY: Peptide
; LOCATION: 1..24
; US-08-238-130-2

Query Match 72.0%; Score 59; DB 1; Length 248;
Best Local Similarity 62.5%; Pred. No. 0.021;
Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEPPFMV 16
Db 25 IVGTSASAGDFFIV 40

Search completed: April 11, 2003, 17:58:50
Job time : 17 secs
```


GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2003, 17:54:35 ; Search time 35 Seconds
(without alignments)
60.915 Million cell updates/sec

Title: US-10-006-223-1
Perfect score: 82
Sequence: 1 VVCGTRAAQGEFFFMV 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
1 number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	78	95.1	15	21	Protease CGW-3 N-t
2	67	81.7	20	21	N-terminus of Stre
3	62	75.6	20	21	N-terminus of nove
4	61	74.4	258	23	Trichoderma harzia
5	59	72.0	248	15	Trypsin-like prote
6	59	72.0	248	16	Trypsin-like Fusar
7	59	72.0	248	17	Pre-pro-trypsin.
8	59	72.0	248	21	Pre-pro-trypsin am
9	56	68.3	360	22	Drosophila melanog
10	54	65.9	251	22	Drosophila melanog

11	54	65.9	294	22	ABB61105
12	54	65.9	296	20	AY27073
13	54	65.9	418	23	ABB06972
14	53	64.6	855	23	AAE23083
15	53	64.6	902	22	AAE98507
16	53	64.6	902	23	AAU77549
17	53	64.6	902	23	AAU80517
18	52	63.4	20	20	AA33343
19	52	63.4	20	21	AA22964
20	52	63.4	20	21	AA93934
21	52	63.4	20	22	AAE07940
22	52	63.4	20	23	AAO21356
23	51	62.2	241	18	AAW22987
24	51	62.2	241	22	AAE06936
25	51	62.2	241	23	AAE22837
26	51	62.2	262	22	AB58183
27	51	62.2	418	23	AB06967
28	51	62.2	620	21	AA43748
29	51	62.2	683	21	AA19551
30	51	62.2	762	21	AA90284
31	51	62.2	851	22	AB11428
32	51	62.2	851	22	AAW25628
33	51	62.2	855	20	AA06671
34	51	62.2	855	21	AA19552
35	51	62.2	855	22	AAE06930
36	51	62.2	855	22	AAE98500
37	51	62.2	855	22	AAE35465
38	51	62.2	932	22	ABG21442
39	50	61.0	25	20	AA33333
40	50	61.0	25	21	AAW22955
41	50	61.0	25	21	AAE93924
42	50	61.0	25	22	AAE07930
43	50	61.0	25	23	AAO21347
44	50	61.0	271	23	ABB62380
45	50	61.0	418	23	ABB06971

ALIGNMENTS

RESULT 1
AAB26847
ID AAB26847 standard; peptide; 15 AA.
AC AAB26847;
XX
XX
DT 29-JAN-2001 (first entry)
XX
DE Protease CGW-3 N-terminal peptide sequence.
XX
KW Protease; CGW-3; plasminogen activator; fibrinolysis;
KW arterial thrombosis.
XX
OS Streptomyces sp.
XX
PN CN1260394-A.
XX
PD 19-JUL-2000.
XX
PF 21-OCT-1999; 99CN-0121864.
XX
PR 21-OCT-1999; 99CN-0121864.
XX
(CHME-) CHINESE ACAD MEDICAL SCI INST BIOTECHNOL.
XX
PI Wang Y, Wu L, Chen F;
XX
DR WPI; 2000-588106/56.
XX
PT New plasmin CGW-3 and its preparation method -
XX
PS Claim 1; Page 1; 8pp; Chinese.
XX

Drosophila melanog
L. bimastus plasmi
Guinea pig AST pro
Epithin protein.
Murine epithin. M
Murine type II mem
Mouse epithelin-li
Crayfish hydrolase
Crayfish protease
N-terminal of a cr
N-terminal of cray
Crayfish protease
Human serine prote
Human membrane-ty
Human matricase o
Drosophila melanog
Porcine AST protei
Human cancer assoc
Human matricase (
Human peptidase, H
Human membrane-ty
Human protein sequ
Tumour antigen der
Human matricase.
Human membrane-ty
Human TAG-15. Ho
Novel human diagno
P. monodon hydrola
Tiger prawn trypti
N-terminal of a sh
N-terminal of P. m
Panaeus monodon tr
Drosophila melanog
Rabbit AST protein

CC The present invention relates to a protease termed CGW-3 which is
 CC obtained by from soil streptomyces strain C3662. CGW-3 is a serine
 CC protease and possesses fibrinolysis activity and can activate
 CC plasminogen. The N-terminal portion of CGW-3 is represented by the
 CC present sequence. The CGW-3 protease can be used in the treatment of
 CC arterial thrombosis.

XX Sequence 15 AA;

SQ Query Match 95.1%; Score 78; DB 21; Length 15;
 Best Local Similarity 100.0%; Pred. No. 1.5e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VVGGTAAQGEFFPMV 15

1 VVGGTAAQGEFFPMV 15

RESULT 2

AAB03088
 AAB03088 standard; peptide; 20 AA.

AC AAB03088;

XX 10-OCT-2000 (first entry)

XX N-terminus of Streptomyces griseus trypsin.

XX Trypsin; bacterial; serine protease; Trichoderma; coagulant;
 KW hypertensive; antiinflammatory; leather preparation; silk treatment.

XX Streptomyces griseus.

XX JP2000116377-A.

XX 25-APR-2000.

XX 08-OCT-1998; 98JP-0303263.

XX 08-OCT-1998; 98JP-0303263.

XX (AMANO) AMANO PHARM KK.

XX WPI; 2000-369402/32.

XX A new serine protease and its preparation, used clinically in blood
 PT coagulation, hypotension and anti-inflammation -

XX Example 3; Page 7; 9pp; Japanese.

CC The invention relates to a novel serine protease from Trichoderma sp.
 CC No.9064. The N-terminus of this protease is given in AAB03085. The novel
 CC protease has trypsin-like activity, specifically cleaving peptide chains
 CC on the carboxyl side of a basic amino acid (e.g., arginine or lysine).
 CC It has a pH optimum of 7-8, and is stable between pH 5 and pH10. It has
 CC a temperature optimum of approximately 40 degrees Celsius, and is stable
 CC between 40 and 50 degrees Celsius. The invention also relates to a
 CC method for the preparation of the novel serine protease, and the use of
 CC the protease in protein degradation. The enzyme has coagulant,
 CC hypertensive and anti-inflammatory effects. It may also be used in the
 CC preparation of leather, for raw silk treatment and for the preparation of
 CC protein hydrolysate. Sequences AAB03086-B03089 represent the N-terminal
 CC 20 amino acids of trypsin from a variety of organisms.

XX Sequence 20 AA;

SQ Query Match 81.7%; Score 67; DB 21; Length 20;
 Best Local Similarity 81.2%; Pred. No. 0.00015;
 Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16

DB 1 VVGGTAAQGEFFPMV 16

RESULT 3

AAB03085

ID AAB03085 standard; peptide; 20 AA.

XX AC AAB03085;

XX 10-OCT-2000 (first entry)

XX N-terminus of novel serine protease from Trichoderma sp. No.9064.

XX Serine protease; bacterial; trypsin activity; coagulant; hypertensive;
 KW antiinflammatory; leather preparation; silk treatment.

XX Trichoderma sp. No.9064.

XX JP2000116377-A.

XX 25-APR-2000.

XX 08-OCT-1998; 98JP-0303263.

XX 08-OCT-1998; 98JP-0303263.

XX (AMANO) AMANO PHARM KK.

XX WPI; 2000-369402/32.

XX A new serine protease and its preparation, used clinically in blood
 PT coagulation, hypotension and anti-inflammation -

XX Example 3; Page 7; 9pp; Japanese.

CC This sequence represents the N-terminal 20 amino acids of a novel serine
 CC protease from Trichoderma sp. No.9064. The novel protease has trypsin-
 CC like activity, specifically cleaving peptide chains on the carboxyl side
 CC of a basic amino acid (e.g., arginine or lysine). It has a pH optimum of
 CC 7-8, and is stable between pH 5 and pH10. It has a temperature optimum of
 CC approximately 40 degrees Celsius, and is stable between 40 and 50 degrees
 CC Celsius. The invention also relates to a method for the preparation of
 CC the novel serine protease, and the use of the protease in protein
 CC degradation. The enzyme has coagulant, hypertensive and anti-inflammatory
 CC effects. It may also be used in the preparation of leather, for raw silk
 CC treatment and for the preparation of protein hydrolysate.

XX Sequence 20 AA;

SQ Query Match 75.6%; Score 62; DB 21; Length 20;
 Best Local Similarity 68.8%; Pred. No. 0.0011;
 Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGGTAAQGEFFPMV 16

DB 1 VVGGTAAQGEFFPMV 16

RESULT 4

ABB83181

ID ABB83181 standard; Protein; 258 AA.

XX AC ABB83181;

XX 16-AUG-2002 (first entry)

XX Trichoderma harzianum Pral protease.

XX Pral; protease; enzyme; fungicide; insecticide; serine-peptidase.

XX Trichoderma harzianum.

XX Key Location/Qualifiers

FT Misc-difference 236

FT XX WO200244359-A1. /note= "Encoded by GAC"

FN XX 06-JUN-2002.

PD XX 30-NOV-2001; 2001WO-ES00471.

PF XX 01-DEC-2000; 2000ES-0002897.

PR XX (NEWB-) NEWBIOTECHNIC SA.

PA (UYSE-) UNIV SEVILLA.

PP (UYSA-) UNIV SALAMANCA.

PI Suarez Fernandez B, Rey Barrera M, Monte Vazquez E;

PN Llobell Gonzalez A;

XX WPI; 2002-471830/50.

DR N-PSDB; ABN83394.

DB New proteolytic enzyme from Trichoderma harzianum, useful e.g. for protecting plants against fungal attack, also related nucleic acid

XX Claim 1; Page 44-46; 51pp; Spanish.

XX The present sequence is the protein sequence for Pral protease from Trichoderma harzianum. Pral, a serine-peptidase, catalyses irreversible inactivation of enzymes and proteins essential for pathogenicity of fungi. Pral can be used to degrade proteins and peptides, especially structural components of the cell walls of fungi, insects and arachnids. Particularly, Pral can be used (optionally in combination with chemical fungicides) to protect plants, animals, harvested crops and foods against fungi; to generate protoplasts and yeast extracts; for recovery of manoproteins; in preparation of wine and (grape) juice; for removal of dental plaque; in tooth-, denture- and contact lens-cleaning solutions; to remove biofilms; to treat or clean textiles; as disinfectant and to prevent contamination of analytical samples.

XX Sequence 258 AA;

Query Match 74.4%; Score 61; DB 23; Length 258;

Best Local Similarity 68.8%; Pred. No. 0.026;

Matches 11; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16

DB :|||||:|||||:

30 IVGTTAAUGGEFFVIV 45

RESULT 5

AAR66999

ID AAR66999 standard; Protein; 248 AA.

XX AAR66999;

12-JUN-1995 (first entry)

Trypsin-like protease.

Trypsin; protease; detergent; surfactant.

Fusarium oxysporum DSM 2672.

Key Location/Qualifiers

Peptide 1..23

/label= Pre-propeptide

WO9425583-A.

10-NOV-1994.

04-MAY-1994; 94WO-DK00177.

PR 05-MAY-1993; 93DK-0000523.

XX (NOVO) NOVO-NORDISK AS.

PA Branner S, Hastrup S;

PI WPI; 1994-358261/44.

DR N-PSDB; AAQ74774.

XX Recombinant trypsin-like protease - useful as e.g. additive enzyme in detergent compsn.

PT Disclosure; Page 33-34; 44pp; English.

XX An active recombinant trypsin-like protease comprises residues 25-224 of the protein given in AAR66999. cDNA encoding the protease was isolated from a cDNA library of F. oxysporum DSM 2682. Expression in Bacillus, Streptomyces, Escherichia coli, Saccharomyces, Aspergillus or Fusarium hosts, with addition of Bacillus metallo protease to the fermentation broth, yielded the mature protease.

XX Sequence 248 AA;

Query Match 72.0%; Score 59; DB 15; Length 248;

Best Local Similarity 62.5%; Pred. No. 0.055;

Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGCTRAAQGEFFPMV 16

DB :|||||:|||||:

25 IVGTTASAGDFPFIV 40

RESULT 6

AAR84716

ID AAR84716 standard; Protein; 248 AA.

XX AAR84716;

19-MAR-1996 (first entry)

Trypsin-like Fusarium protease.

Trypsin-like protease; detergent; surfactant; enzyme engineering; enzyme stabilisation.

Fusarium oxysporum DSM 2672.

Key Location/Qualifiers

Peptide 1..6

/label= Pro-peptide

Peptide 8..24

/label= Sig_peptide

Region 32..36

/label= Loop-I

Region 44..47

/label= Loop-II

Region 55..58

/label= Loop-III

Region 67..75

/label= Loop-IV

Region 81..88

/label= Loop-V

Region 99..105

/label= Loop-VI

Region 116..124

/label= Loop-VII

Region 130..137

/label= Loop-VIII

Region 151..158

/label= Loop-IX

Region 170..186

/label= Loop-X

Region 191..196

The invention relates to a non-toxic, non-pathogenic recombinant Fusarium host cell of the section *discolor*, with ATCC accession number 20334. The

TO TRY TO GET THE

CC cell is used in the recombinant production of proteins. The present
 CC sequence represents the F. oxysporum pre-pro-trypsin protein sequence,
 CC which is an example of an enzyme which can be expressed by the cells of
 CC the invention. The cells are useful for expressing heterologous proteins
 CC especially fungal enzymes such as alkaline endoglucanase or alkaline
 CC proteases, e.g. F. oxysporum pre-pro trypsin gene, and also hormones,
 CC growth factors and receptors. The cells are non-toxic and are efficient
 CC in the recombinant production of fungal enzymes.

XX SQ Sequence 248 AA;

Query Match 72.0%; Score 59; DB 21; Length 248;
 Best Local Similarity 62.5%; Pred. No. 0.055;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 Db 25 IVGTSASAGDFPFIV 40

LT 9
 ABB60104
 ID ABB60104 standard; Protein; 360 AA.
 XX
 AC ABB60104;

26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 7104.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.
 OS WO200171042-A2.
 RN
 XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.
 N-PSDB; ABL04207.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 7104; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI16175), and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 360 AA;

Query Match 68.3%; Score 56; DB 22; Length 360;
 Best Local Similarity 66.7%; Pred. No. 0.27;

Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VVGTRAAQGEFFPM 15
 :|||:|:|:|:
 Db 102 IVGTRASGKEFFPM 116

RESULT 10
 ABB64578
 ID ABB64578 standard; Protein; 251 AA.

XX ABB64578;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 20526.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
 DR N-PSDB; ABL08681.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 20526; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
 CC sequences (ABLI01840-ABLI16175), and the encoded proteins
 CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 251 AA;

Query Match 65.9%; Score 54; DB 22; Length 251;
 Best Local Similarity 50.0%; Pred. No. 0.39;
 Matches 8; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 :|||:|:|:|:
 Db 28 IVGTRAREGQPHQI 43

RESULT 11
 ABB61105
 ID ABB61105 standard; Protein; 294 AA.

XX ABB61105;

XX

DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 10107.
 DE
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.

 Venter JC, Adams M, Li PWD, Myers EW;

 WPI; 2001-656860/75.
 N-PSDB; ABL05208.
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 10107; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

 Sequence 294 AA;

 Query Match 65.9%; Score 54; DB 22; Length 294;
 Best Local Similarity 66.7%; Pred. No. 0.46;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

 Db 1 VVGGTAAAGGFFPFM 15
 :|||:|||||:
 37 IVGAKAAGREFFPM 51

 RESULT 12
 AAY27073
 ID AAY27073 standard; Protein; 296 AA.
 XX
 AC AAY27073;
 XX
 DT 18-OCT-1999 (first entry)
 XX
 DE L. bimastus plasmin mature protein sequence.
 XX
 KW Lumbricus bimastus; earthworm; plasmin gene.
 XX
 OS Lumbricus bimastus.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 4
 FT Misc-difference /note= "encoded by GC"
 FT Misc-difference 12
 FT Misc-difference /note= "encoded by TGC"
 FT

FT Misc-difference 243
 /note= "encoded by TAG"
 FT Misc-difference 268
 /note= "encoded by TAA"
 FT Misc-difference 274
 /note= "encoded by TAA"
 FT Misc-difference 279
 /note= "encoded by TAA"
 FT Misc-difference 280
 /note= "encoded by TAA"
 FT Misc-difference 292
 /note= "encoded by TAA"
 FT Misc-difference 296
 /note= "encoded by TAG"
 XX
 PN CN1208770-A.
 XX
 PD 24-FEB-1999.
 XX
 XX 11-JUN-1998; 98CN-0102257.
 XX
 PR 11-JUN-1998; 98CN-0102257.
 PA (VIRO-) VIROLOGY RES INST CHINA PREVENTIVE MEDIC.
 XX
 PI Fu S, Liang G, Meng X;
 XX
 DR WPI; 1999-313741/27.
 DR N-PSDB; AAX89869.
 XX
 XX Lumbrical fibrinolysin gene nucleotide series, and method for clone
 PT of same - is composed of 88 nucleotides in which position 1-726
 PT nucleotide is gene matured peptide sequence, position 72-729
 PT nucleotide is terminal codon TAG
 XX
 PS Claim 3; Page 1; 9pp; Chinese.
 XX
 CC The invention provides a Lumbricus bimastus (a kind of earthworm) plasmin
 CC gene. The gene is composed of 88 nucleotides in which position 1-726
 CC nucleotide is gene matured peptide sequence, position 72-729 nucleotide
 CC is terminal codon TAG. The expressed protein from the gene possesses
 CC tremendous medical prospect. The present sequence represents the
 CC L. bimastus plasmin gene mature protein sequence.
 XX
 SQ Sequence 296 AA;

 Query Match 65.9%; Score 54; DB 20; Length 296;
 Best Local Similarity 56.2%; Pred. No. 0.47;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

 QY 1 VVGGTAAAGGFFPFMV 16
 :|||:|||||:
 Db 1 VVGGTAAAGGFFPWL 16

 RESULT 13
 ABB06972
 ID ABB06972 standard; Protein; 418 AA.
 XX
 AC ABB06972;
 XX
 DT 20-JUN-2002 (first entry)
 XX
 DE Guinea pig AST protein sequence SEQ ID NO:37.
 XX
 KW Airway specific trypsin-like protease; AST; protease activated receptor;
 KW PAR; mucus production promotion; cell proliferation; calcium flow;
 KW EGFR pathway activation; epidermal growth factor receptor; enzyme.
 XX
 OS Cavia porcellus.
 XX
 XX Key Location/Qualifiers
 FT Key
 FT Peptide 1..186

FT Protein /label= propeptide
 FT 187..418
 XX /label= trypsin-like protein
 PN WC000218562-A1.
 XX
 XX
 XX PD 07-MAR-2002.
 XX
 XX PF 28-AUG-2001; 2001WO-JP07349.
 XX
 XX 28-AUG-2000; 2000JP-0257104.
 PR 05-MAR-2001; 2001JP-0059753.
 XX
 XX (TEIJ) TEIJIN LTD.
 PA

XX Eguchi H, Chokki M, Yamamura S, Mita R, Masegi T;
 XX WPI; 2002-315539/35.
 DR N-PSDB; ABL50743.
 DR

PT Airway-specific trypsin-like enzymes for use in diagnosis and screening
 XX compounds or polypeptides as inhibitors of AST activity, PAR activation
 XX and mucus production, and judging therapeutic efficacy
 PS Example 10; Page 146-148; 165pp; Japanese.

CC The present invention describes an airway-specific trypsin-like enzyme
 CC (AST) is a protein comprising the whole or a part of the amino acid
 CC sequence of (AB06964 (I)), in which a propeptide moiety containing all
 CC or part of an amino acid sequence of AST between Met at position 1 and
 CC Arg at position 186 is bonded to a trypsin-like moiety containing Ile at
 CC position 187 to Ile at position 418 in a 533 amino acid sequence via a
 CC disulfide bond. The AST enzyme can be used in the diagnosis and screening
 CC of compounds and polypeptides as inhibitors of AST activity, protease
 CC activated receptor (PAR) activation, mucus production promotion, cell
 CC proliferation, calcium flow into cells or EGFR pathway activation by AST
 CC and judging therapeutic efficacy. The present sequence represents Guinea
 CC pig AST from the present invention.
 XX
 XX Sequence 418 AA;

Query Match 65.9%; Score 54; DB 23; Length 418;
 Best Local Similarity 62.5%; Pred. No. 0.69;
 Matches 10; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 |||||:|:|:|:
 187 VVGTTQADQGDWPMQV 202

RESULT 14
 AAE23083
 ID AAE23083 standard; Protein; 855 AA.

XX AAE23083;

DT 21-AUG-2002 (first entry)

XX Epithin protein.

XX Transgenic; transgenic animal; pharmacological therapy; gene therapy;
 XX phenotype modulation; genetic disease; epithin protein.

XX Unidentified.

XX WO200203787-A2.

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21427.

XX 06-JUL-2000; 2000US-216109P.

PR 06-JUL-2000; 2000US-216251P.

PR 06-JUL-2000; 2000US-216258P.
 PR 06-JUL-2000; 2000US-216768P.
 PR 10-JUL-2000; 2000US-217449P.
 PR 10-JUL-2000; 2000US-217450P.
 PR 10-JUL-2000; 2000US-217660P.
 PR 27-JUL-2000; 2000US-221491P.
 PR 27-JUL-2000; 2000US-221669P.
 PR 27-JUL-2000; 2000US-221670P.
 PR 07-AUG-2000; 2000US-223170P.
 PR 07-AUG-2000; 2000US-223172P.
 PR 07-AUG-2000; 2000US-223460P.
 PR 26-OCT-2000; 2000US-244037P.
 PR 26-OCT-2000; 2000US-244111P.
 PR 26-JUN-2001; 2001US-301217P.
 XX
 XX (DELT-) DELTAGEN INC.

XX Allen KD, Leviten MW;

PI
 XX WPI; 2002-154853/20.
 DR N-PSDB; AAD37039.

XX Novel non-human transgenic animal, preferably transgenic mice
 XX comprising disruption in target gene, e.g., trypsin gene, useful for
 XX identifying an agent that modulates expression or function of target
 XX gene
 XX Disclosure; Fig 6; 74pp; English.

PS The present invention relates to non-human transgenic animals preferably
 XX transgenic mice comprising disruption in target gene such as trypsin,
 XX gene. The invention also relates to compositions and methods relating
 XX to the characterization of gene functions. The transgenic animals are
 XX useful for identifying an agent that modulates the expression or function
 XX of a target. They are useful for identifying an agent that modulates a
 XX phenotype associated with a disruption in trypsin genes or limulus
 XX clotting factor protease-like genes by administering an agent to the
 XX transgenic animal and determining whether the agent modulates the
 XX phenotype where the agent has effect on decreased body weight, decreased
 XX thymus weight, decreased thymus to body weight ratio, increased pre-pulse
 XX inhibition, significant decrease in their response latency to the hot
 XX plate test or a decreased response threshold to metrazol. Agents that
 XX modulate the expression, function or activity of the target gene are
 XX useful for treating a disorder associated with a mutation in trypsin
 XX gene or in limulus clotting factor protease-like gene. The transgenic
 XX animals are useful for testing the efficacy of proposed genetic and
 XX pharmacological therapies for human genetic diseases. They are useful
 XX as models for diseases, disorders or conditions associated with
 XX phenotypes relating to a disruption in a target and to identify drugs,
 XX pharmaceuticals, therapies and interventions which may be effective in
 XX treating a disease or other phenotypic characteristics of the animal.
 XX The present sequence is, epithin protein. This sequence is used in the
 XX exemplification of the invention.

XX Sequence 855 AA;

Query Match 64.6%; Score 53; DB 23; Length 855;
 Best Local Similarity 62.5%; Pred. No. 2.2;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVGTRAAQGEFFPMV 16
 |||||:|:|:|:
 615 VVGGINADEGEWPMQV 630

Db

RESULT 15

AAE23083

ID AAB98507 standard; Protein; 902 AA.

XX AAB98507;

XX 03-AUG-2001 (first entry)

DT

DE	Murine epithin.
KW	Murine; TAGD-15; cytostatic; vaccine; ovarian tumour; cancer; epithin;
KW	tumour antigen-derived gene 15; serine protease.
XX	
OS	Mus musculus.
PN	WO200129056-A1.
PD	26-APR-2001.
XX	
XX	20-OCT-2000; 2000WO-US29095.
XX	20-OCT-1999; 99US-0421213.
XX	(UYAR-) UNIV ARKANSAS.
PT	O'Brien TJ, Tanimoto H;
PI	WPI; 2001-381031/40.
XX	
PT	Novel extracellular serine protease, termed tumor antigen-derived gene
PT	15 protein overexpressed in carcinomas and DNA encoding it, for
PT	diagnosis, treatment, prevention of cancer, particularly breast,
XX	ovarian cancer -
PS	Disclosure; Page 97-99; 130pp; English.
XX	
CC	The present invention relates to human tumour antigen-derived gene 15
CC	(TAGD-15) protein and coding sequence (see AAH23601 and AAB95800).
CC	TAGD-15 is an extracellular serine protease. It was found that TAGD-15 is
CC	over-expressed in ovarian tumours. TAGD-15 protein or its fragments of
CC	9-20 residues that lack TAGD-15 protease activity are useful for
CC	vaccinating an individual against TAGD-15, having, suspected of having or
CC	at risk of getting cancer. Furthermore, the TAGD-15 gene can be used as a
CC	diagnostic or therapeutic target in cancer. The present sequence was used
XX	in a sequence homology alignment with TAGD-15.
XX	
SQ	Sequence 902 AA;
	Query Match 64.6%; Score 53; DB 22; Length 902;
	st Local Similarity 62.5%; Pred. No. 2.4;
	tches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Qy	1 VVGTRAAQGPFPMV 16 : :
Db	615 VVGGTNADEGWPMOV 630 : :